

“Mankiw & Taylor’s *Economics* is a superb book for all students approaching this subject for the first time. The book is both intuitive, with plenty of examples enabling students to link economic theory and facts, and rigorous, with analytical supplements and extensive exercises allowing students to go into depth if they wish to. This book will make students love this subject and it is simply excellent.”

Dr Gaia Garino, Principal Teaching Fellow in Economics, University of Leicester, UK

“A very well written and modern text, covering a wide and exhaustive range of topics to be taught in introduction to economics classes. The accessible language and approach is ideal for all students including those to whom English is a second language, and a key pedagogical strength of the book is the many examples which show students how to apply key economics topics within their everyday lives.”

Prof. Erich Ruppert, Faculty of Economics, Business and Law, University of Aschaffenburg, Germany

Now firmly established as one of the leading economics principles texts in the UK and Europe, this exciting, new third edition by N. Gregory Mankiw (Harvard University) and Mark P. Taylor (Warwick University), has undergone some significant restructuring and reorganization to more directly match economics students’ course structures and learning and assessment needs. There are new sections covering microeconomic and macroeconomic topics and concepts in more depth, whilst at the same time retaining the book’s reputation for clarity, authority and real world relevance.

New to the third edition:

- > In direct response to user feedback, the chapter structure has been reorganized to better map to typical UK and European course structures
- > Brand new Case Studies, In the News features, and examples throughout
- > Supplementary maths content provided separately
- > Enhanced lecturer and student digital support resources including articles from The Economist with associated discussion questions linked to every chapter and new assessment practice questions which can be used for tutorial discussion and assignments
- > New coverage on information and behavioural economics, business cycles, supply-side policies the role of empirical evidence, and economic rent
- > Fully updated to reflect the economic arguments which have surfaced following the financial crisis
- > Fresh, modern text design with accessible layout and approach

About the Authors

N. Gregory Mankiw, Professor of Economics, Harvard University, USA

Mark P. Taylor, Professor of Economics and Dean of Warwick Business School, University of Warwick, UK

Economics is essential reading for all students taking introductory economics modules on undergraduate courses and within an economics component of postgraduate and MBA courses.



For your lifelong learning solutions, visit www.cengage.co.uk
Purchase your next print book, e-book or e-chapter at www.cengagebrain.com

ISBN 978-1-4080-9379-5



9 781408 093795 >

Economics



Mankiw
Taylor



N. Gregory Mankiw **Mark P. Taylor**

Economics



$$\sum_T^M 3^e$$

N. Gregory Mankiw **Mark P. Taylor**

Economics



Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

Economics, 3rd Edition**N. Gregory Mankiw and Mark P. Taylor**

Publisher: Andrew Ashwin

Commissioning Editor: Annabel Ainscow

Senior Production Editor: Alison Burt

Senior Manufacturing Buyer: Eyvett Davis

Typesetter: diacriTech

Cover design: Adam Renvoize

Text design: Design Deluxe

© 2014, Cengage Learning EMEA

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, or applicable copyright law of another jurisdiction, without the prior written permission of the publisher.

While the publisher has taken all reasonable care in the preparation of this book, the publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions from the book or the consequences thereof.

Products and services that are referred to in this book may be either trademarks and/or registered trademarks of their respective owners. The publishers and author/s make no claim to these trademarks. The publisher does not endorse, and accepts no responsibility or liability for, incorrect or defamatory content contained in hyperlinked material.

For product information and technology assistance,
contact emea.info@cengage.com

For permission to use material from this text or product,
and for permission queries,
email emea.permissions@cengage.com.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 978-1-4080-9379-5

Cengage Learning EMEACheriton House, North Way, Andover, Hampshire SP10 5BE
United KingdomCengage Learning products are represented in Canada by
Nelson Education Ltd.For your lifelong learning solutions, visit
www.cengage.co.ukPurchase your next print book, e-book or e-chapter at
www.cengagebrain.com

BRIEF CONTENTS

About the authors ix
Preface x
Supplements xi
Acknowledgements xiii

PART 1 Introduction to economics 1

- 1 Ten principles of economics 1
- 2 Thinking like an economist 17

PART 2 Supply and demand: How markets work 41

- 3 The market forces of supply and demand 41
- 4 Elasticity and its applications 72
- 5 Background to demand: The theory of consumer choice 102
- 6 Background to supply: Firms in competitive markets 134

PART 3 Markets, efficiency and welfare 169

- 7 Consumers, producers and the efficiency of markets 169
- 8 Supply, demand and government policies 187

PART 4 The economics of the public sector 203

- 9 The tax system and the costs of taxation 203

PART 5 Inefficient market allocations 221

- 10 Public goods, common resources and merit goods 221
- 11 Externalities and market failure 239
- 12 Information and behavioural economics 264

PART 6 Firm behaviour and market structures 279

- 13 Firms' production decisions 279
- 14 Market structures I: Monopoly 290
- 15 Market structures II: Monopolistic competition 314
- 16 Market structures III: Oligopoly 329

PART 7 Factor markets 355

- 17 The economics of labour markets 355

PART 8 Inequality 385

- 18 Income inequality and poverty 385

PART 9 Trade 405

- 19 Interdependence and the gains from trade 405

PART 10 The data of macroeconomics 437

- 20 Measuring a nation's income 437
- 21 Measuring the cost of living 456

PART 11 The real economy in the long run 473

- 22 Production and growth 473
- 23 Unemployment 497

PART 12 Interest rates, money and prices in the long run 519

- 24 Saving, investment and the financial system 519
- 25 The basic tools of finance 539
- 26 The monetary system 558
- 27 Money growth and inflation 583

PART 13 The macroeconomics of open economies 605

- 28 Open-economy macroeconomics: Basic concepts 605
- 29 A macroeconomic theory of the open economy 622

PART 14 Short-run economic fluctuations 637

- 30 Business cycles 637
- 31 Keynesian economics and IS-LM analysis 655
- 32 Aggregate demand and aggregate supply 679
- 33 The influence of monetary and fiscal policy on aggregate demand 702
- 34 The short-run trade-off between inflation and unemployment 721
- 35 Supply-side policies 745

PART 15 International macroeconomics 759

- 36 Common currency areas and European monetary union 759
- 37 The financial crisis and sovereign debt 782

CONTENTS

About the authors ix
Preface x
Supplements xi
Acknowledgements xiii

PART 1 INTRODUCTION TO ECONOMICS 1

1 Ten principles of economics 1

What is economics? 1
How people make decisions 2
How people interact 6
How the economy as a whole works 8
Conclusion 12

2 Thinking like an economist 17

Introduction 17
The economist as scientist 17
The economist as policy advisor 23
Why economists disagree 24
Let's get going 28
Appendix
Graphing and the tools of economics: A brief review 30

PART 2 SUPPLY AND DEMAND: HOW MARKETS WORK 41

3 The market forces of supply and demand 41

Markets and competition 41
Demand 43
Supply 50
Supply and demand together 56
Conclusion: How prices allocate resources 67

4 Elasticity and its applications 72

The price elasticity of demand 72
Other demand elasticities 81
Price elasticity of supply 83
Applications of supply and demand elasticity 95

5 Background to demand: The theory of consumer choice 102

The standard economic model 102
The budget constraint: What the consumer can afford 104
Preferences: What the consumer wants 108
Optimization: What the consumer chooses 113
Summary: Do people really think this way? 126
Behavioural approaches to consumer behaviour 126

6 Background to supply: Firms in competitive markets 134

The costs of production 134
Production and costs 135
The various measures of cost 138
Costs in the short run and in the long run 145
Summary 146
Returns to scale 147
What is a competitive market? 150
Profit maximization and the competitive firm's supply curve 153
The supply curve in a competitive market 160
Conclusion: Behind the supply curve 164

PART 3 MARKETS, EFFICIENCY AND WELFARE 169

7 Consumers, producers and the efficiency of markets 169

Consumer surplus 170
Producer surplus 176
Market efficiency 179
Conclusion: Market efficiency and market failure 183

8 Supply, demand and government policies 187

- Controls on prices 187
- Taxes 192
- Subsidies 198
- Conclusion 199

**PART 4
THE ECONOMICS OF THE
PUBLIC SECTOR 203****9 The tax system and the costs of taxation** 203

- Taxes and efficiency 203
- The deadweight loss of taxation 204
- Administrative burden 210
- The design of the tax system 211
- Taxes and equity 214
- Conclusion 216

**PART 5
INEFFICIENT MARKET
ALLOCATIONS 221****10 Public goods, common resources and merit goods** 221

- The different kinds of goods 222
- Public goods 223
- Common resources 230
- Merit goods 232
- Conclusion 234

11 Externalities and market failure 239

- Externalities 239
- Externalities and market inefficiency 240
- Private solutions to externalities 244
- Public policies towards externalities 248
- Public/private policies towards externalities 251
- Government failure 254
- Conclusion 260

12 Information and behavioural economics 264

- Principal and agent 264
- Asymmetric information 265
- Deviations from the standard economic model 271
- Conclusion 274

**PART 6
FIRM BEHAVIOUR AND MARKET
STRUCTURES 279****13 Firms' production decisions** 279

- Isoquants and isocosts 279
- The least-cost input combination 284
- Conclusion 286

14 Market structures I: Monopoly 290

- Imperfect competition 290
- Why monopolies arise 291
- How monopolies make production and pricing decisions 294
- The welfare cost of monopoly 300
- Price discrimination 303
- Public policy towards monopolies 307
- Conclusion: The prevalence of monopoly 309

15 Market structures II: Monopolistic competition 314

- Competition with differentiated products 315
- Advertising and branding 319
- Contestable markets 323
- Conclusion 324

16 Market structures III: Oligopoly 329

- Characteristics of oligopoly 329
- Game theory and the economics of cooperation 335
- Models of oligopoly 344
- Public policy toward oligopolies 347
- Conclusion 350

PART 7 **FACTOR MARKETS 355**

17 The economics of labour markets 355

- The demand for labour 356
- The supply of labour 360
- Equilibrium in the labour market 364
- Wage differentials 367
- The economics of discrimination 373
- The other factors of production: Land and capital 376
- Economic rent 379
- Conclusion 380

PART 8 **INEQUALITY 385**

18 Income inequality and poverty 385

- The measurement of inequality 386
- The political philosophy of redistributing income 395
- Conclusion 401

PART 9 **TRADE 405**

19 Interdependence and the gains from trade 405

- The production possibilities frontier 405
- International trade 410
- The principle of comparative advantage 414
- The determinants of trade 418
- The winners and losers from trade 420
- Restrictions on trade 423
- Conclusion 431

PART 10 **THE DATA OF MACROECONOMICS 437**

20 Measuring a nation's income 437

- The economy's income and expenditure 438
- The measurement of gross domestic product 439
- The components of GDP 441
- Real versus nominal GDP 445
- GDP and economic well-being 449
- Conclusion 452

21 Measuring the cost of living 456

- The consumer prices index 456
- Correcting economic variables for the effects of inflation 465
- Conclusion 468

PART 11 **THE REAL ECONOMY IN THE LONG RUN 473**

22 Production and growth 473

- Economic growth around the world 474
- Growth theory 476
- Productivity: Its role and determinants 477
- Economic growth and public policy 482
- Conclusion: The importance of long-run growth 493

23 Unemployment 497

- Identifying unemployment 497
- Job search 504
- Minimum wage laws 506
- Unions and collective bargaining 508
- The theory of efficiency wages 510
- The cost of unemployment 511
- Conclusion 514

PART 12

INTEREST RATES, MONEY AND PRICES IN THE LONG RUN 519

24 Saving, investment and the financial system 519

Financial institutions in the economy 520
 Saving and investment in the national income accounts 527
 The market for loanable funds 530
 Conclusion 535

25 The basic tools of finance 539

Present value: Measuring the time value of money 539
 Managing risk 541
 Asset valuation 548
 Conclusion 554

26 The monetary system 558

The meaning of money 559
 The role of central banks 565
 The European Central Bank and the Eurosystem 566
 The Bank of England 567
 Banks and the money supply 568
 Conclusion 579

27 Money growth and inflation 583

The classical theory of inflation 584
 The costs of inflation 594
 Conclusion 599

PART 13

THE MACROECONOMICS OF OPEN ECONOMIES 605

28 Open-economy macroeconomics: Basic concepts 605

The international flows of goods and capital 606
 The prices for international transactions: Real and nominal exchange rates 609
 A first theory of exchange rate determination: Purchasing power parity 612
 Conclusion 618

29 A macroeconomic theory of the open economy 622

Supply and demand for loanable funds and for foreign currency exchange 622
 Equilibrium in the open economy 625
 How policies and events affect an open economy 628
 Conclusion 634

PART 14

SHORT-RUN ECONOMIC FLUCTUATIONS 637

30 Business cycles 637

Trend growth rates 638
 Causes of changes in the business cycle 645
 Business cycle models 646
 Conclusion 650

31 Keynesian economics and IS-LM analysis 655

The Keynesian cross 655
 The multiplier effect 659
 The IS and LM curves 665
 General equilibrium using the IS-LM model 667
 From IS-LM to aggregate demand 669
 Conclusion 675

32 Aggregate demand and aggregate supply 679

Three key facts about economic fluctuations 679
 Explaining short-run economic fluctuations 680
 The aggregate demand curve 682
 The aggregate supply curve 686
 Two causes of economic fluctuations 691
 New Keynesian economics 697
 Conclusion 698

33 The influence of monetary and fiscal policy on aggregate demand 702

How monetary policy influences aggregate demand 702
 How fiscal policy influences aggregate demand 709
 Using policy to stabilize the economy 713
 Conclusion 716

34 The short-run trade-off between inflation and unemployment 721

- The Phillips curve 721
- Shifts in the Phillips curve: The role of expectations 724
- The long-run vertical Phillips curve as an argument for Central Bank independence 730
- Shifts in the Phillips curve: the role of supply shocks 732
- The cost of reducing inflation 734
- Inflation targeting 739
- Conclusion 740

35 Supply-side policies 745

- Shifts in the aggregate supply curve 745
- Types of supply-side policies 749
- Conclusion 756

**PART 15
INTERNATIONAL
MACROECONOMICS 759**

36 Common currency areas and European monetary union 759

- The euro 760
- The single European market and the euro 760
- The benefits and costs of a common currency 762
- The theory of optimum currency areas 765
- Is Europe an optimum currency area? 768
- Fiscal policy and common currency areas 772
- Conclusion 777

37 The financial crisis and sovereign debt 782

- Bubbles and speculation 782
- The sovereign debt crisis 793
- Austerity policies – too far too quickly? 797

- Glossary 805
- Index 814
- Credits 820

ABOUT THE AUTHORS

AUTHORS

N. GREGORY MANKIW is Professor of Economics at Harvard University. As a student, he studied economics at Princeton University and the Massachusetts Institute of Technology (MIT). As a teacher he has taught macroeconomics, microeconomics, statistics and principles of economics. Professor Mankiw is a prolific writer and a regular participant in academic and policy debates. In addition to his teaching, research and writing, Professor Mankiw has been a research associate of the National Bureau of Economic Research, an advisor to the Federal Reserve Bank of Boston and the Congressional Budget Office. From 2003 to 2005, he served as chairman of the US President's Council of Economic Advisors and was an advisor to Presidential candidate Mitt Romney during the 2012 US presidential election. Professor Mankiw lives in Wellesley, Massachusetts, with his wife Deborah, their three children and their border terrier Tobin.

MARK P. TAYLOR is Dean of Warwick Business School at the University of Warwick and Professor of International Finance. He obtained his first degree in philosophy, politics and economics from Oxford University and his Master's degree in economics from London University, from where he also holds a doctorate in economics and international finance. Professor Taylor has taught economics and finance at various universities (including Oxford, Warwick and New York) and at various levels (including principles courses, advanced undergraduate and advanced postgraduate courses). He has also worked as a senior economist at the International Monetary Fund and at the Bank of England and, before becoming Dean of Warwick Business School, was a managing director at BlackRock, the world's largest financial asset manager, where he worked on international asset allocation based on macroeconomic analysis. His research has been extensively published in scholarly journals and he is today one of the most highly cited economists in the world. Professor Taylor lives with his family in a 15th-century farmhouse near Stratford upon Avon, Warwickshire, where he collects clocks and keeps bees.

CONTRIBUTING AUTHOR

ANDREW ASHWIN has over 20 years experience as a teacher of economics. He has an MBA and is currently researching for a PhD investigating assessment and the notion of threshold concepts in economics. Andrew is an experienced author, writing a number of texts for students at different levels and journal publications related to his PhD research, and learning materials for the website Biz/ed, which was based at the University of Bristol. Andrew was Chair of Examiners for a major awarding body for business and economics in England and is a consultant for the UK regulator, Ofqual. Andrew has a keen interest in assessment and learning in economics and has received accreditation as a Chartered Assessor with the Chartered Institute of Educational Assessors. He is also Editor of the Economics, Business and Enterprise Association (EBEA) journal. Andrew lives in Rutland with his wife Sue and their twin sons Alex and Johnny.

PREFACE

The third edition of *Economics* has a different look to the previous two editions. Feedback from users, both students and instructors, has resulted in some reorganization of the material and some new sections covering more depth in both micro- and macroeconomic issues. Readers should note that this edition adapts Greg Mankiw's best-selling US undergraduate *Economics* text to reflect the needs of students and instructors in the UK and European market. As each new edition is written, the adaptation evolves and develops an identity distinct from the original US edition on which it is based.

We have tried to retain the lively, engaging writing style and to continue to have the novice economics student in mind. Economics touches every aspect of our lives and the fundamental concepts which are introduced can be applied across a whole range of life experiences. 'Economics is a study of mankind in the ordinary business of life.' So wrote Alfred Marshall, the great 19th-century British economist, in his textbook, *Principles of Economics*. As you work through the contents of this book you would be well advised to remember this.

Whilst the news might focus on the world of banking and finance, tax and government policy, economics provides much more than a window on these worlds. It provides an understanding of decision making and the process of decision making across so many different aspects of life. You may be considering travelling abroad, for example, and are shocked at the price you have to pay for injections against tropical diseases. Should you decide to try and do without the injections? Whilst the amount of money you are expected to give up seems high, it is a small price to pay when you consider the trade-off – the potential cost to you and your family of contracting a serious disease. This is as much economics as monetary policy decisions about interest rates and firm's decisions on investment.

Welcome to the wonderful world of economics – learn to think like an economist and a whole new world will open up to you.

Maths for Mankiw Taylor Economics is available for purchase as a supplementary resource carefully explaining and teaching the maths concepts and formulae underlying many of the key chapter topics.

DIGITAL SUPPORT RESOURCES

Dedicated Instructor Resources

To discover the dedicated instructor online support resources accompanying this textbook, instructors should register here for access:

<http://login.cengage.com>

Resources include:

- Instructor's Manual
- Testbank
- PowerPoint slides



Instructor access

Instructors can access CourseMate by registering at <http://login.cengage.com> or by speaking to their local Cengage Learning EMEA representative.

Instructor resources

Instructors can use the integrated Engagement Tracker in CourseMate to track students' preparation and engagement. The tracking tool can be used to monitor progress of the class as a whole, or for individual students.

Student access

Log In & Learn In 4 Easy Steps

1. To register a product using the access code printed on the inside front-cover of the book please go to: <http://login.cengagebrain.com>
2. Register as a new user or log in as an existing user if you already have an account with Cengage Learning or CengageBrain.com
3. Follow the online prompts
4. If your instructor has provided you with a course key, you will be prompted to enter this after opening your digital purchase from your CengageBrain account homepage.

Student resources

CourseMate offers a range of interactive learning tools tailored to the third edition of Mankiw & Taylor, *Economics* including:

- Multiple choice and self-test questions
- Interactive eBook

- Critical Thinking Questions
- Case Studies
- Discussion questions
- Links to useful websites



Instructor access

Instructors can access Aplia by registering at <http://login.cengage.com> or by speaking to their local Cengage Learning EMEA representative.

Instructor resources

Cengage Learning EMEA's Aplia is a fully tailored online learning and assessment solution which offers an easy-to-use course management system, to save instructors valuable time they'd otherwise spend on routine assignment setting and marking. To date, Aplia has been used by more than 1 000 000 students at over 1 300 institutions.

Student access

On the recommendation of their instructor, students can purchase Aplia by searching for *ECONOMICS* on www.cengagebrain.co.uk

Student resources

Aplia is dedicated to improving learning by increasing student effort and engagement, and provides detailed explanations for every question to help students stay focused, alert and thinking critically. The core elements of an Aplia course include:

- Subject-relevant experiments
- Embedded eBook
- An easy-to-use course management system
- Personalised customer support
- Interactive, automatically graded problem sets
- Chapter assignments
- Tutorials
- News analyses

ACKNOWLEDGEMENTS

Michael Barrow, *University of Sussex, UK*
Brian Bell, *London School of Economics, UK*
Thomas Braeuninger, *University of Mannheim, Germany*
Eleanor Denny, *Trinity College Dublin, UK*
Gaia Garino, *University of Leicester, UK*
Chris Grammenos, *American College of Thessaloniki, Greece*
Getinet Haile, *University of Nottingham, UK*
Luc Hens, *Vrije Uni, Belgium*
William Jackson, *University of York, UK*
Colin Jennings, *King's College London, UK*
Sarah Louise Jewell, *University of Reading, UK*
Arie Kroon, *Utrecht Hogeschool, The Netherlands*

Jassodra Maharaj, *University of East London, UK*
Paul Melessen, *Hogeschool van Amsterdam, The Netherlands*
Jørn Rattsø, *Norwegian University of Science & Technology, Norway*
Frédéric Robert-Nicoud, *University of Geneva, Switzerland*
Jack Rogers, *University of Exeter, UK*
Erich Ruppert, *Hochschule Aschaffenburg, Germany*
Noel Russell, *University of Manchester, UK*
Munacinga Simatele, *University of Hertfordshire, UK*
Robert Simmons, *University of Lancaster, UK*
Alison Sinclair, *University of Nottingham, UK*

PART 1

INTRODUCTION TO

ECONOMICS

1

TEN PRINCIPLES OF

ECONOMICS

WHAT IS ECONOMICS?

The word *economy* comes from the Greek word *oikonomos*, which means ‘one who manages a household’. At first, this origin might seem peculiar. But, in fact, households and economies have much in common.

A household faces many decisions. It must decide which members of the household do which tasks and what each member gets in return: Who cooks dinner? Who does the laundry? Who gets the extra slice of cake at tea time? Who chooses what TV programme to watch? In short, the household must allocate its scarce resources among its various members, taking into account each member’s abilities, efforts and desires.

Like a household, a society faces many decisions. A society must decide what jobs will be done and who will do them. It needs some people to grow food, other people to make clothing and still others to design computer software. Once society has allocated people (as well as land, buildings and machines) to various jobs, it must also allocate the output of goods and services that they produce. It must decide who will eat caviar and who will eat potatoes. It must decide who will drive a Mercedes and who will take the bus.

The Economic Problem

These decisions can be summarized as representing the economic problem. There are three questions that any society has to face:

- What goods and services should be produced?
- How should these goods and services be produced?
- Who should get the goods and services that have been produced?

The answer to these questions would be simple if resources were so plentiful that society could produce everything any of its citizens could ever want, but this is not the case. Society will never have enough

resources to produce the goods and services which will satisfy the wants and needs of its citizens. These resources can be broadly classified into three categories:

- **Land** – all the natural resources of the earth. This includes things like mineral deposits such as iron ore, gold and copper, fish in the sea, coal and all the food products that land yields. David Ricardo (1817) in his book *On the Principles of Political Economy and Taxation* referred to land as the ‘original and indestructible powers of the soil’.
- **Labour** – the human effort both mental and physical that goes in to production. A worker in a factory producing precision tools, an investment banker, a road sweeper, a teacher – these are all forms of labour.
- **Capital** – the equipment and structures used to produce goods and services. Capital goods include machinery in factories, buildings, tractors, computers, cooking ovens – anything where the good is not used for its own sake but for the contribution it makes to production.

land all the natural resources of the earth

labour the human effort both mental and physical that goes in to production

capital the equipment and structures used to produce goods and services

Scarcity and Choice

What resources society does have need to be managed. The management of society’s resources is important because resources are scarce. **Scarcity** means that society has limited resources and therefore cannot produce all the goods and services people wish to have. Just as a household cannot give every member everything he or she wants, a society cannot give every individual the highest standard of living to which he or she might aspire.

Economics is the study of how society manages its scarce resources and attempts to answer the three key questions we noted above. In most societies, resources are allocated through the combined actions of millions of households and firms through a system of markets. Economists:

- Study how people make decisions: how much they work, what they buy, how much they save and how they invest their savings.
- Study how people interact with one another. For instance, they examine how the multitude of buyers and sellers of a good together determine the price at which the good is sold and the quantity that is sold.
- Analyse forces and trends that affect the economy as a whole, including the growth in average income, the fraction of the population that cannot find work and the rate at which prices are rising.

scarcity the limited nature of society’s resources

economics the study of how society manages its scarce resources

Although the study of economics has many facets, the field is unified by several central ideas. In the rest of this chapter we look at the *Ten Principles of Economics*. Don’t worry if you don’t understand them all at first, or if you don’t find them completely convincing. In the coming chapters we will explore these ideas more fully. The ten principles are introduced here just to give you an overview of what economics is all about. You can think of this chapter as a ‘preview of coming attractions’.

HOW PEOPLE MAKE DECISIONS

There is no mystery to what an ‘economy’ is. Whether we are talking about the economy of a group of countries such as the European Union (EU), or the economy of one particular country, such as India, or of the whole world, an economy is just a group of people interacting with one another as they go about their lives. **The economy** refers to all the production and exchange activities that take place every day – all the buying and selling. The level of **economic activity** is how much buying and selling goes on in the economy over a period of time.

the economy all the production and exchange activities that take place every day
economic activity how much buying and selling goes on in the economy over a period of time

Because the behaviour of an economy reflects the behaviour of the individuals who make up the economy, we start our study of economics with four principles of individual decision making.

Principle 1: People Face Trade-offs

The first lesson about making decisions is summarized in an adage popular with economists: 'There is no such thing as a free lunch.' To get one thing that we like, we usually have to give up another thing that we also like. Making decisions requires trading off the benefits of one goal against those of another.

Consider a student who must decide how to allocate her most valuable resource – her time. She can spend all of her time studying economics which will bring benefits such as a better class of degree; she can spend all her time enjoying leisure activities which yield different benefits; or she can divide her time between the two. For every hour she studies, she gives up an hour she could have devoted to spending time in the gym, riding a bicycle, watching TV, napping or working at her part-time job for some extra spending money.

Consider parents deciding how to spend their family income. They can buy food, clothing or a family holiday. Or they can save some of the family income for retirement or perhaps to help the children buy a house or a flat when they are grown up. When they choose to spend an extra euro on one of these goods, they have one less euro to spend on some other good.

When people are grouped into societies, they face different kinds of trade-offs. The classic trade-off is between spending on defence and spending on food. The more we spend on national defence to protect our country from foreign aggressors, the less we can spend on consumer goods to raise our standard of living at home. Also important in modern society is the trade-off between a clean environment and a high level of income. Laws that require firms to reduce pollution raise the cost of producing goods and services. Because of the higher costs, these firms end up earning smaller profits, paying lower wages, charging higher prices, or some combination of these three. Thus, while pollution regulations give us the benefit of a cleaner environment and the improved levels of health that come with it, they have the cost of reducing the incomes of the firms' owners, workers and customers.

Another trade-off society faces is between efficiency and equity. Efficiency means that society is getting the most it can (depending how this is defined) from its scarce resources. **Equity** means that the benefits of those resources are distributed fairly among society's members. In other words, efficiency refers to the size of the economic cake, and equity refers to how the cake is divided. Often, when government policies are being designed, these two goals conflict.

equity – the property of distributing economic prosperity fairly among the members of society

Consider, for instance, policies aimed at achieving a more equal distribution of economic well-being. Some of these policies, such as the social security system or unemployment insurance, try to help those members of society who are most in need. Others, such as income tax, ask the financially successful to contribute more than others to support government spending. Although these policies have the benefit of achieving greater equity, they have a cost in terms of reduced efficiency. When the government redistributes income from the rich to the poor, it reduces the reward for working hard; as a result, people work less and produce fewer goods and services. In other words, when the government tries to cut the economic cake into more equal slices, the cake gets smaller.

Recognizing that people face trade-offs does not by itself tell us what decisions they will or should make. A student should not abandon the study of economics just because doing so would increase the time available for leisure. Society should not stop protecting the environment just because environmental regulations reduce our material standard of living. The poor should not be ignored just because helping

them distorts work incentives. Nevertheless, acknowledging life's trade-offs is important because people are likely to make good decisions only if they understand the options that they have available.

A typical supermarket shelf offering a variety of cereals. What are the trade-offs an individual faces in this situation?



SELF TEST Does the adage 'there is no such thing as a free lunch' simply refer to the fact that someone has to have paid for the lunch to be provided and served? Or does the recipient of the 'free lunch' also incur a cost?

Principle 2: The Cost of Something is What You Give Up to Get It

Because people face trade-offs, making decisions requires comparing the costs and benefits of alternative courses of action. In many cases, however, the cost of some action is not as obvious as it might first appear.

Consider, for example, the decision whether to go to university. The benefit is intellectual enrichment and a lifetime of better job opportunities. But what is the cost? To answer this question, you might be tempted to add up the money you spend on tuition fees, books, room and board. Yet this total does not truly represent what you give up to spend a year at university.

The first problem with this answer is that it includes some things that are not really costs of going to university. Even if you decided to leave full-time education, you would still need a place to sleep and food to eat. Room and board are part of the costs of higher education only to the extent that they might be more expensive at university than elsewhere. Indeed, the cost of room and board at your university might be less than the rent and food expenses that you would pay living on your own. In this case, the savings on room and board are actually a benefit of going to university.

The second problem with this calculation of costs is that it ignores the largest cost of a university education – your time. When you spend a year listening to lectures, reading textbooks and writing essays, you cannot spend that time working at a job. For most students, the wages given up to attend university are the largest single cost of their higher education.

The **opportunity cost** of an item is what you give up to get that item. When making any decision, such as whether to go to university, decision makers should be aware of the opportunity costs that accompany each possible action. In fact, they usually are. University-age rugby, basketball or golf players who can earn large sums of money if they opt out of higher education and play professional sport are well aware that their opportunity cost of going to university is very high. It is not surprising that they often decide that the benefit is not worth the cost.

opportunity cost – whatever must be given up to obtain some item; the value of the benefits foregone (sacrificed)

SELF TEST Assume the following costs are incurred by a student over a three-year course at a university:

- Tuition fees at €9,000 per year = €27,000
- Accommodation, based on an average cost of €4,500 a year = €13,500
- Opportunity cost based on average earnings foregone of €15,000 per year = €45,000
- Total cost = €85,500

● Given this relatively large cost why does anyone want to go to university?

Principle 3: Rational People Think at the Margin

Decisions in life are rarely straightforward and usually involve problems. At dinner time, the decision you face is not between fasting or eating as much as you can, but whether to take that extra serving of pizza. When examinations roll around, your decision is not between completely failing them or studying 24 hours a day, but whether to spend an extra hour revising your notes instead of watching TV. Economists use the term **marginal changes** to describe small incremental adjustments to an existing plan of action. Keep in mind that 'margin' means 'edge', so marginal changes are adjustments around the edges of what you are doing.

marginal changes small incremental adjustments to a plan of action

In many situations, people make the best decisions by thinking at the margin. Suppose, for instance, that you were considering whether to study for a Master's degree having completed your undergraduate studies. To make this decision, you need to know the additional benefits that an extra year in education would offer (higher wages throughout your life and the sheer joy of learning) and the additional costs that you would incur (another year of tuition fees and another year of foregone wages). By comparing these *marginal benefits* and *marginal costs*, you can evaluate whether the extra year is worthwhile.

Individuals and firms can make better decisions by thinking at the margin. A rational decision maker takes an action if and only if the marginal benefit of the action exceeds the marginal cost.

Principle 4: People Respond to Incentives

Because people make decisions by comparing costs and benefits, their behaviour may change when the costs or benefits change. That is, people respond to incentives. When the price of an apple rises, for instance, people decide to eat more pears and fewer apples because the cost of buying an apple is higher. At the same time, apple farmers decide to hire more workers and harvest more apples, because the benefit of selling an apple is also higher. As we shall see, the effect of price on the behaviour of buyers and sellers in a market – in this case, the market for apples – is crucial for understanding how the economy works.

Public policymakers should never forget about incentives, because many policies change the costs or benefits that people face and, therefore, alter behaviour. A tax on petrol, for instance, encourages people to drive smaller, more fuel efficient cars. It also encourages people to switch and use public transport rather than drive, or to move closer to where they work. When policymakers fail to consider how their policies affect incentives, they often end up with results they did not intend. For example, the UK government provided tax relief on business premises that were not being used as an incentive to the owners to find new uses or owners for the buildings. The government decided to remove the tax relief and suggested that in doing so there would now be an incentive for owners of premises to get them back into use as quickly as possible so that they avoided losing the tax relief. Unfortunately, as the new policy came into being the economy was going through a severe recession. It was not easy for owners of premises to find new tenants let alone get new businesses created in these empty properties. Some property owners decided that rather than have to pay tax on these properties it was cheaper to demolish them. Is this the outcome the government wanted? Almost certainly not.

This is an example of the general principle that people respond to incentives. Many incentives that economists study are straightforward and others more complex. No one is surprised, for example, that people might switch to driving smaller cars where petrol taxes and thus the price of fuel is relatively high.

Yet, as the example of the removal of tax allowances on empty business premises shows, policies can have effects that are not obvious in advance. When analysing any policy, we must consider not only the direct effects but also the indirect effects that work through incentives. If the policy changes incentives, it will cause people to alter their behaviour.

SELF TEST Many people across the EU are without work and claiming benefits. Governments throughout the EU are trying to cut spending but find themselves having to spend more on welfare benefits for the unemployed. What sort of incentives might governments put in place to encourage workers off welfare and into work? What might be the unintended consequences of the incentives you identify?

HOW PEOPLE INTERACT

The first four principles discussed how individuals make decisions. As we go about our lives, many of our decisions affect not only ourselves but other people as well. The next three principles concern how people interact with one another.

Principle 5: Trade Can Make Everyone Better Off

America and China are competitors to Europe in the world economy. In some ways this is true, because American and Chinese firms produce many of the same goods as European firms. Toy manufacturers compete for the same customers in the market for toys. Fruit farmers compete for the same customers in the market for fruit.

Yet it is easy to be misled when thinking about competition among countries. Trade between Europe and the United States and China is not like a sports contest, where one side wins and the other side loses (a zero-sum game). In fact, the opposite is true: trade between two economies can make each economy better off.

To see why, consider how trade affects your family. When a member of your family looks for a job, he or she competes against members of other families who are looking for jobs. Families also compete against one another when they go shopping, because each family wants to buy the best goods at the lowest prices. So, in a sense, each family in the economy is competing with all other families.

Despite this competition, your family would not be better off isolating itself from all other families. If it did, your family would need to grow its own food, make its own clothes and build its own home. Clearly, your family gains much from its ability to trade with others. Trade allows each person to specialize in the activities he or she does best, whether it is farming, sewing or home building. By trading with others, people can buy a greater variety of goods and services at lower cost.

Countries as well as families benefit from the ability to trade with one another. Trade allows countries to specialize in what they do best and to enjoy a greater variety of goods and services. The Japanese and the Americans, as well as the Koreans and the Brazilians, are as much Europe's partners in the world economy as they are competitors.

Principle 6: Markets Are Usually a Good Way to Organize Economic Activity

The collapse of communism in the Soviet Union and Eastern Europe in the 1980s may be the most important change in the world during the past half century. Communist countries worked on the premise that central planners in the government were in the best position to guide economic activity and answer the three key questions of the economic problem. These planners decided what goods and services were produced, how much was produced, and who produced and consumed these goods and services. The theory behind central planning was that only the government could organize economic activity in a way that promoted economic well-being for the country as a whole.

Today, most countries that once had centrally planned economies such as Russia, Poland, Angola, Mozambique and the Democratic Republic of Congo have abandoned this system and are trying to develop

market economies. In a **market economy**, the decisions of a central planner are replaced by the decisions of millions of firms and households. Firms decide whom to hire and what to make. Households decide which firms to work for and what to buy with their incomes. These firms and households interact in the marketplace, where prices and self-interest guide their decisions.

market economy an economy that addresses the three key questions of the economic problem through allocating resources through the decentralized decisions of many firms and households as they interact in markets for goods and services

At first glance, the success of market economies is puzzling. After all, in a market economy, no one is considering the economic well-being of society as a whole. Free markets contain many buyers and sellers of numerous goods and services, and all of them are interested primarily in their own well-being. Yet, despite decentralized decision making and self-interested decision makers, market economies have proven remarkably successful in organizing economic activity in a way that promotes overall economic well-being.

FYI

Adam Smith and the Invisible Hand

Adam Smith's great work *An Inquiry into the Nature and Causes of the Wealth of Nations* was published in 1776 and is a landmark in economics. In its emphasis on the invisible hand of the market economy, it reflected a point of view that was typical of so-called 'enlightenment' writers at the end of the 18th century – that individuals are usually best left to their own devices, without government guiding their actions. This political philosophy provides the intellectual basis for the market economy.

Why do decentralized market economies work so well? Is it because people can be counted on to treat one another with love and kindness? Not at all. Here is Adam Smith's description of how people interact in a market economy:

Man has almost constant occasion for the help of his brethren, and it is vain for him to expect it from their benevolence only. He will be more

likely to prevail if he can interest their self-love in his favour, and show them that it is for their own advantage to do for him what he requires of them. ... It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. ...

Every individual ... neither intends to promote the public interest, nor knows how much he is promoting it. ... He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

Smith is saying that participants in the economy are motivated by self-interest and that the 'invisible hand' of the marketplace guides this

self-interest into promoting general economic well-being.

Many of Smith's insights remain at the centre of modern economics. Our analysis in the coming chapters will allow us to express Smith's conclusions more precisely and to analyse fully the strengths and weaknesses of the market's invisible hand.



One of our goals in this book is to understand how Smith's invisible hand works its magic. As you study economics, you will learn that prices are the instrument with which the invisible hand directs economic activity. Prices reflect both the value of a good to society and the cost to society of making the good. Because households and firms look at prices when deciding what to buy and sell, they unknowingly take into account the social benefits and costs of their actions. As a result, prices guide these individual decision makers to reach outcomes that, in many cases, maximize the welfare of society as a whole.

Principle 7: Governments Can Sometimes Improve Market Outcomes

If the invisible hand of the market is so wonderful, why do we need government? One answer is that the invisible hand needs government to protect it. Markets work only if property rights are enforced. A farmer won't grow food if he expects his crop to be stolen, and a restaurant won't serve meals unless it is assured that customers will pay before they leave. We all rely on government provided police and courts to enforce our rights over the things we produce.

Yet there is another answer to why we need government: although markets are usually a good way to organize economic activity, this rule has some important exceptions. There are two broad reasons for a government to intervene in the economy – to promote efficiency and to promote equity. That is, most policies aim either to enlarge the economic cake or to change the way in which the cake is divided.

Although the invisible hand often leads markets to allocate resources efficiently, that is not always the case. Economists use the term **market failure** to refer to a situation in which the market on its own fails to produce an efficient allocation of resources. One possible cause of market failure is an **externality**, which is the uncompensated impact of one person's actions on the well-being of a bystander (a third party). For instance, the classic example of an external cost is pollution. Another possible cause of market failure is **market power**, which refers to the ability of a single person or business (or group of businesses) to unduly influence market prices. In the presence of market failure, well designed public policy can enhance economic efficiency.

market failure a situation where scarce resources are not allocated to their most efficient use
externality the cost or benefit of one person's decision on the well-being of a bystander (a third party) which the decision maker does not take into account in making the decision
market power the ability of a single economic agent (or small group of agents) to have a substantial influence on market prices

The invisible hand may also fail to ensure that economic prosperity is distributed equitably. One of the three questions society has to address is who gets what is produced? A market economy rewards people according to their ability to produce things for which other people are willing to pay. The world's best footballer earns more than the world's best chess player simply because people are willing to pay more to watch football than chess. That individual is getting more of what is produced as a result of his earnings. The invisible hand does not ensure that everyone has sufficient food, decent clothing and adequate health care. Many public policies, such as income tax and the social security system, aim to achieve a more equitable distribution of economic well-being.

To say that the government *can* improve on market outcomes at times does not mean that it always *will*. Public policy is made not by angels but by a political process that is far from perfect. Sometimes policies are designed simply to reward the politically powerful. Sometimes they are made by well-intentioned leaders who are not fully informed. One goal of the study of economics is to help you judge when a government policy is justifiable to promote efficiency or equity, and when it is not.

HOW THE ECONOMY AS A WHOLE WORKS

We started by discussing how individuals make decisions and then looked at how people interact with one another. All these decisions and interactions together make up 'the economy'. The last three of our ten principles concern the workings of the economy as a whole.

Microeconomics and Macroeconomics

Economics is studied on various levels. The first seven principles involve the study of the decisions of individual households and firms and the interaction of households and firms in markets for specific goods and services. In the last three principles we are looking at the operation of the economy as a whole, which is just the sum of the activities of all these decision makers in all these markets.

Since roughly the 1930s, the field of economics has traditionally been divided into two broad subfields.

Microeconomics is the study of how households and firms make decisions and how they interact in specific markets. **Macroeconomics** is the study of economy-wide phenomena. A microeconomist might study the effects of a congestion tax on the use of cars in a city centre, the impact of foreign competition on the European car industry or the effects of attending university on a person's lifetime earnings. A macroeconomist might study the effects of borrowing by national governments, the changes over time in an economy's rate of unemployment or alternative policies to raise growth in national living standards.

microeconomics the study of how households and firms make decisions and how they interact in markets
macroeconomics the study of economy-wide phenomena, including inflation, unemployment and economic growth

Microeconomics and macroeconomics are closely intertwined. Because changes in the overall economy arise from the decisions of millions of individuals, it is impossible to understand macroeconomic developments without considering the associated microeconomic decisions. For example, a macroeconomist might study the effect of a cut in income tax on the overall production of goods and services in an economy. To analyse this issue, he or she must consider how the tax cut affects the decisions of households concerning how much to spend on goods and services.

Despite the inherent link between microeconomics and macroeconomics, the two fields are distinct. In economics, it may seem natural to begin with the smallest unit and build up. Yet doing so is neither necessary nor always the best way to proceed. Because microeconomics and macroeconomics address different questions, they sometimes take quite different approaches and are often taught in separate courses.

A key concept in macroeconomics is **economic growth** – the percentage increase in the number of goods and services produced in an economy over a period of time, usually expressed over a quarter and annually.

economic growth the increase in the amount of goods and services in an economy over a period of time

Principle 8: An Economy's Standard of Living Depends on its Ability to Produce Goods and Services

Table 1.1 shows **gross domestic product per capita (head)** of the population in a number of selected countries expressed in U.S. dollars. It is clear that many of the advanced economies have a relatively high income per capita; in Norway it is an enviable \$98,102, the Netherlands \$50,087 and Germany \$43,689.

gross domestic product per capita (head) the market value of all goods and services produced within a country in a given period of time divided by the population of a country to give a per capita figure

Moving away from the prosperous economies of Western Europe, we begin to see differences in income and living standards around the world that are quite staggering. For example, average income in Yemen was \$1,361 whilst in Afghanistan average income is just over a half a per cent of the size of per-capita income in Norway.

TABLE 1.1

Gross Domestic Product Per Capita, Current Prices US dollars 2011

| | |
|--------------------|--------|
| Afghanistan | 576 |
| Austria | 49,707 |
| Belgium | 46,469 |
| Benin | 802 |
| Bolivia | 1,421 |
| China | 5,445 |
| Estonia | 16,556 |
| Finland | 49,391 |
| Germany | 43,689 |
| Hungary | 21,732 |
| Italy | 36,116 |
| Japan | 45,903 |
| Kenya | 808 |
| Netherlands | 50,087 |
| Norway | 98,102 |
| Portugal | 22,330 |
| Russian Federation | 13,089 |
| Spain | 32,244 |
| Sweden | 56,927 |
| Switzerland | 80,391 |
| Turkey | 10,498 |
| United Kingdom | 38,818 |
| United States | 48,442 |
| Yemen | 1,361 |

Not surprisingly, this large variation in average income is reflected in various other measures of the quality of life and **standard of living**. Citizens of high-income countries have better nutrition, better health care and longer life expectancy than citizens of low-income countries, as well as more TV sets, more gadgets and more cars.

standard of living refers to the amount of goods and services that can be purchased by the population of a country. Usually measured by the inflation-adjusted (real) income per head of the population

Changes in the standard of living over time are also large. Over the last 5 years, economic growth in Albania has grown at about 4.68 per cent per year, in China at about 10.5 per cent a year but in Latvia the economy has shrunk by around 1.4 per cent over the same time period (Source: World Bank).

What explains these large differences in living standards among countries and over time? The answer is surprisingly simple. Almost all variation in living standards is attributable to differences in countries' **productivity** – that is, the amount of goods and services produced from each hour of a worker's time. In nations where workers can produce a large quantity of goods and services per unit of time, most people enjoy a high standard of living; in nations where workers are less productive, most people must endure a more meagre existence. Similarly, the growth rate of a nation's productivity determines the growth rate of its average income.

productivity the quantity of goods and services produced from each hour of a worker or factor of production's time

The fundamental relationship between productivity and living standards is simple, but its implications are far-reaching. If productivity is the primary determinant of living standards, other explanations must be of secondary importance. For example, it might be tempting to credit trade unions or minimum wage laws for the rise in living standards of workers over the past 50 years. Yet the real hero of workers is their rising productivity.

The relationship between productivity and living standards also has profound implications for public policy. When thinking about how any policy will affect living standards, the key question is how it will affect our ability to produce goods and services. To boost living standards, policymakers need to raise productivity by ensuring that workers are well educated, have the tools needed to produce goods and services, and have access to the best available technology.

Principle 9: Prices Rise When the Government Prints Too Much Money

In Zimbabwe in March 2007 inflation was reported to be running at 2,200 per cent. That meant that a good priced at the equivalent of Z\$2.99 in March 2006 would be priced at Z\$65.78 just a year later. In February 2008, inflation was estimated at 165,000 per cent. Five months later it was reported as 2,200,000 per cent. In July 2008 the government issued a Z\$100 billion note. At that time it was just about enough to buy a loaf of bread. Estimates for inflation in Zimbabwe in July 2008 put the rate of growth of prices at 231,000,000 per cent. In January 2009, the government issued Z\$10, 20, 50 and 100 trillion dollar notes – 100 trillion is 100 followed by 12 zeros. This episode is one of history's most spectacular examples of **inflation**, an increase in the overall level of prices in the economy.

inflation an increase in the overall level of prices in the economy

High inflation is a problem because it imposes various costs on society; keeping inflation at a low level is a goal of economic policymakers around the world. What causes inflation? In almost all cases of high or persistent inflation, the culprit turns out to be the same – growth in the quantity of money. When a government creates large quantities of the nation's money, the value of the money falls. As outlined above, the Zimbabwean government was issuing money at ever higher denominations. It is generally accepted that there is a relationship between the growth in the quantity of money and the rate of growth of prices.

Principle 10: Society Faces a Short-run Trade-off Between Inflation and Unemployment

When the government increases the amount of money in the economy, one result is inflation. Another result, at least in the short run, is a lower level of unemployment. The curve that illustrates this short-run trade-off between inflation and unemployment is called the **Phillips curve**, after the economist who first examined this relationship while working at the London School of Economics.

Phillips curve a curve that shows the short run trade-off between inflation and unemployment

The Phillips curve remains a controversial topic among economists, but most economists today accept the idea that society faces a short-run trade-off between inflation and unemployment. This simply means that, over a period of a year or two, many economic policies push inflation and unemployment in opposite directions. Policymakers face this trade-off regardless of whether inflation and unemployment both start out at high levels at low levels or somewhere in-between.

The trade-off between inflation and unemployment is only temporary, but it can last for several years. The Phillips curve is, therefore, crucial for understanding many developments in the economy. In particular, it is important for understanding the **business cycle** – the irregular and largely unpredictable fluctuations in economic activity, as measured by the number of people employed or the production of goods and services.

business cycle fluctuations in economic activity such as employment and production

Policymakers can exploit the short-run trade-off between inflation and unemployment using various policy instruments. By changing the amount that the government spends, the amount it taxes and the amount of money it prints, policymakers can influence the combination of inflation and unemployment that the economy experiences. Because these instruments of monetary and fiscal policy are potentially so powerful, how policymakers should use these instruments to control the economy, if at all, is a subject of continuing debate.

SELF TEST What is the difference between microeconomics and macroeconomics? Write down three questions that the study of microeconomics might be concerned with and three questions that might be involved in the study of macroeconomics

CONCLUSION

You now have a taste of what economics is all about. In the coming chapters we will develop many specific insights about people, markets and economies. Mastering these insights will take some effort, but it is not an overwhelming task. The field of economics is based on a few basic ideas that can be applied in many different situations.

Throughout this book we will refer back to the *Ten Principles of Economics* highlighted in this chapter and summarized in Table 1.2 which can be seen as building blocks for your study of the subject; you should keep these building blocks in mind. Even the most sophisticated economic analysis is built using the ten principles introduced here.

TABLE 1.2 Ten Principles of Economics

| | |
|----------------------------------|--|
| How people make decisions | 1. People face trade-offs 2. The cost of something is what you give up to get it 3. Rational people think at the margin |
| How people interact | 4. People respond to incentives 5. Trade can make everyone better off 6. Markets are usually a good way to organize economic activity 7. Governments can sometimes improve market outcomes |
| How the economy as a whole works | 8. A country's standard of living depends on its ability to produce goods and services 9. Prices rise when the government prints too much money 10. Society faces a short-run trade-off between inflation and unemployment |

IN THE NEWS



Latest Thinking in Economics – Incentives

One of the *Ten Principles of Economics* is that people respond to incentives. This should not be an entirely surprising principle and may seem like an example of economists making common sense sound more complex. However, the reality is that the complex nature of human beings does make the introduction and effect of incentives much

more challenging than might at first appear.

Gneezy et al. highlight some of these issues (Gneezy, U. Meier, S. and Ray-Biel, P. (2011) 'When and why incentives (don't) work to modify behaviour'. In *Journal of Economic Perspectives*. 25:4, 191–210). They point out that incentives may work better in certain circumstances than

in others and policymakers need to consider a wide variety of issues when deciding on putting incentives in place.

First of all, they have to consider the type of behaviour to be changed. For example, society might want to encourage its citizens to do more, what Gneezy et al. call 'prosocial' behaviour such as donating blood, sperm or organs, increasing the

amount of waste put out for recycling, attending school, college or university, working harder in education to improve grades, improving the environment such as installing insulation or solar panels in homes to reduce energy waste, or finding ways of encouraging people to stop smoking.

Second, we have to consider the parties involved. This can be expressed as a principal-agent issue. The principal is a person or group for whom another person or group, the agent, is performing some act. In encouraging people to stop smoking, the smoker is the agent and society is the principal. Next, we have to consider the type of incentive offered – often this will be monetary. Monetary incentives have two main types of effects which Gneezy et al. refer to as the direct price effect and the psychological effect. Once the behaviour has been identified, the type of incentive and who the principal and agent are, the next question is to consider how the incentive is framed.

At first the solution might be seen as being simple – provide a monetary incentive; pay people to achieve the desired behaviour. The question is, will the incentive work? Gneezy et al. point to a number of reasons why the outcome might not be as obvious as first hoped. They suggest that in some cases, offering monetary incentives can ‘crowd out’ the desired behaviour. Offering a monetary incentive to go to school, donate blood or install solar panels might not have the desired effect. The reasons might be that offering a monetary incentive changes the perceptions of agents. People have intrinsic motivations – personal reasons for particular behaviours. Other people also have perceptions about the behaviour of others, for example someone who donates blood might be seen by others as being ‘nice’. Social norms may also be affected, for example attitudes to recycling of waste or smoking.

Providing a monetary incentive on these behaviours might not necessarily lead to more blood being donated, more recycling and solar panels or less smoking. Gneezy et al. suggest that the reasons may be that monetizing behaviour in this way changes the psychology and the psychology effect can be greater than the direct price effect. The price effect would suggest that if you pay someone to donate more blood, you should get more people donating blood. The reality might be that such incentives reduce blood donorship. Why? People who donate blood might do so out of a personal conviction – they have intrinsic motivations. By offering monetary incentives, the perception of the donor and others might change so that they are not seen as being ‘nice’ any more but as being ‘mercenary’ and not motivated intrinsically but by extrinsic reward – greed, in other words. If the psychological effect outweighs the direct money effect the result could be a reduction in the number of donors.

In the case of cutting smoking, the size of the money effect might be a factor. Principal 5 of *The Ten Principles of Economics* states that rational people think at the margin. With smoking, the marginal decision to have one more cigarette imposes costs and benefits on the smoker – the benefit is the pleasure people get from smoking, the cost the (estimated) 11 minutes of their life that is cut as a result. The problem is that the marginal cost is not tangible at that time and is likely to be outweighed by the marginal benefit (not to mention the addictive qualities of tobacco products). Over time, however, the total benefit of stopping smoking

becomes much greater than the total cost. The incentive offered, therefore, has to be such that it takes into account these marginal decisions and it might be difficult to estimate the size of the incentive needed.

Other issues relating to incentives involve the trust between the principal and agent. If an incentive is provided, for example, then this sends a message that the desired behaviour is not taking place and that there may be a reason for this. This might be that the desired behaviour is not attractive and/or is difficult to carry out. Incentives also send out a message that the principal does not trust the agent’s intrinsic motivation, for example that people will not voluntarily give blood or recycle waste effectively. Some incentives may work to achieve the desired behaviour in the short-term but will this lead to the desired behaviour continuing in the long-term when the incentive is removed?

Finally, incentives might be affected by the way they are framed – how the wording or the benefits of the incentive is presented to the agent by the principal. Gneezy et al. use a very interesting example of this. Imagine a situation, they say, where you meet a person and develop a relationship. You want to provide that person with the incentive to have sex. The effect of the way

Providing a monetary incentive on these behaviours might not necessarily lead to more blood being donated



the incentive is framed might have a considerable effect on the outcome. If, for example, you framed your 'offer' by saying 'I would like to make love to you and to incentivize you to do so I will offer you €50', you might get a very different response to that if you framed it by saying: 'I would like to make love to you – I have bought you a bunch of red roses' (the roses just happened to cost €50).

Finally, the cost effectiveness of incentives has to be considered. Health authorities spend millions of euros across Europe on drugs to reduce blood pressure and cholesterol. Getting people to take more exercise will also help achieve the

same result. What would be more cost-effective and a more efficient allocation of resources? Providing incentives (assuming they work) to encourage people to exercise more by, for example, paying for gym membership, or spending that same money on drugs but not dealing with some of the underlying causes?

Questions

- 1 **Why should people need incentives to do 'good' things like donating blood or putting out more rubbish for recycling?**
- 2 **What is meant by the 'principal-agent' issue?**
- 3 **What might be the price and psychological effect if students were given a monetary incentive to attain top grades in their university exams?**
- 4 **Why might the size of a monetary incentive be an important factor in encouraging desired behaviour and what side-effects might arise if the size of an incentive was increased?**
- 5 **What is 'framing' and why might it be important in the way in which an incentive works? Refer to the need to increase the number of organ donors in your answer to this question.**

FYI



How To Read This Book

Economics is fun, but it can also be hard to learn. Our aim in writing this text has been to make it as easy and as much fun as possible. But you, the student, also have a role to play. Experience shows that if you are actively involved as you study this book, you will enjoy a better outcome, both in your exams and in the years that follow. Here are a few tips about how best to read this book.

1. *Summarize, don't highlight.* Running a yellow marker over the text is too passive an activity to keep your mind engaged. Instead, when you come to the end of a section, take a minute and summarize what you have just learnt in your own words, writing your summary in a note book or on your computer. When you've finished the chapter, compare your summary with the one at the end of the chapter. Did you pick up the main points?
2. *Test yourself.* Throughout the book, the Self Test features offer the chance to test your understanding of the subject matter. Take the opportunity to jot down your ideas and thoughts to the Self Test questions. The tests are meant to assess your basic comprehension and application of the ideas and concepts in the chapter. If you aren't sure your answer is right, you probably need to review the section.
3. *Practise, practise, practise.* At the end of each chapter, Questions for Review test your understanding, and Problems and Applications ask you to apply and extend the material. Perhaps your lecturer will assign some of these exercises as work for seminars and tutorials. If so, do them. If not, do them anyway. The more you use your new knowledge, the more solid it becomes.
4. *Study in groups.* After you've read the book and worked through the problems on your own, get together with other students to discuss the material. You will learn from each other – an example of the gains from trade.
5. *Don't forget the real world.* In the midst of all the numbers, graphs and strange new words, it is easy to lose sight of what economics is all about. The Case Studies and In the News boxes sprinkled throughout this book should help remind you. Don't skip them. They show how the theory is tied to events happening in all of our lives and the questions provided with the In the News features will help you think about issues that you have covered in the chapter and also to apply your understanding to specific contexts. As with the Self Test questions, attempt an answer to the questions to help build your understanding.

SUMMARY

- The fundamental lessons about individual decision making are that people face trade-offs among alternative goals, that the cost of any action is measured in terms of foregone opportunities, that rational people make decisions by comparing marginal costs and marginal benefits, and that people change their behaviour in response to the incentives they face.
- The fundamental lessons about interactions among people are that trade can be mutually beneficial, that markets are usually a good way of coordinating trade among people, and that the government can potentially improve market outcomes if there is some market failure or if the market outcome is inequitable.
- The field of economics is divided into two sub-fields: microeconomics and macroeconomics. Microeconomists study decision making by households and firms and the interaction among households and firms in the marketplace. Macroeconomists study the forces and trends that affect the economy as a whole.
- The fundamental lessons about the economy as a whole are that productivity is the ultimate source of living standards, that money growth is the ultimate source of inflation, and that society faces a short-run trade-off between inflation and unemployment.

QUESTIONS FOR REVIEW

- 1 Give three examples of important trade-offs that you face in your life.
- 2 What is the opportunity cost of going to a restaurant for a meal?
- 3 Water is necessary for life. Is the marginal benefit of a glass of water large or small?
- 4 Why should policymakers think about incentives?
- 5 Why isn't trade among countries like a game, with some winners and some losers?
- 6 What does the 'invisible hand' of the marketplace do?
- 7 Explain the two main causes of market failure and give an example of each.
- 8 What are the two subfields into which economics is divided? Explain what each subfield studies.
- 9 Why is productivity important?
- 10 How are inflation and unemployment related in the short run?

PROBLEMS AND APPLICATIONS

- 1 Describe some of the trade-offs faced by each of the following.
 - a. A family deciding whether to buy a new car.
 - b. A member of the government deciding how much to spend on building a new motorway connecting two main cities.
 - c. A company chief executive officer deciding whether to recommend the acquisition of a smaller firm.
 - d. A university lecturer deciding how much to prepare for her lecture.
- 2 You are trying to decide whether to take a holiday. Most of the costs of the holiday (airfare, hotel, foregone wages) are measured in euros, but the benefits of the holiday are psychological. How can you compare the benefits to the costs?
- 3 You were planning to spend an evening working at your part-time job, but a friend asks you to go to a night club. What is the true cost of going to the night club? Now suppose that you had been planning to spend the evening studying in the library. What is the cost of going to the night club in this case? Explain.
- 4 You win €10,000 on the EuroMillions lottery draw. You have a choice between spending the money now or putting it away for a year in a bank account that pays 5 per cent interest. What is the opportunity cost of spending the €10,000 now?
- 5 The company that you manage has invested €5 million in developing a new product, but the development is not quite finished. At a recent meeting, your sales people report that the introduction of competing products has reduced the expected sales of your new product to €3 million. If it would cost €1 million to finish development and make the product, should you go ahead and do so? What is the most that you should pay to complete development?
- 6 Three managers of the van Heerven Coach Company are discussing a possible increase in production. Each suggests a way to make this decision.

FIRST MANAGER: We need to decide how many additional coaches to produce. Personally, I think we should examine whether our company's productivity – number of

coaches produced per worker per hour – would rise or fall if we increased output.

SECOND MANAGER: We should examine whether our average cost per worker – would rise or fall.

THIRD MANAGER: We should examine whether the extra revenue from selling the additional coaches would be greater or smaller than the extra costs.

Who do you think is right? Why?

- 7 Assume a social security system in a country provides income for people over the age of 65. If a recipient decides to work and earn some income, the amount he or she receives in social security benefits is typically reduced.
 - a. How does the provision of this grant affect people's incentive to save while working?
 - b. How does the reduction in benefits associated with higher earnings affect people's incentive to work past the age of 65?
- 8 Your flatmate is a better cook than you are, but you can clean more quickly than your flatmate can. If your

flatmate did all of the cooking and you did all of the cleaning, would your household chores take you more or less time than if you divided each task evenly? Give a similar example of how specialization and trade can make two countries both better off.

- 9 Explain whether each of the following government activities is motivated by a concern about equity or a concern about efficiency. In the case of efficiency, discuss the type of market failure involved.
 - a. Regulating water prices.
 - b. Regulating electricity prices.
 - c. Providing some poor people with vouchers that can be used to buy food.
 - d. Prohibiting smoking in public places.
 - e. Imposing higher personal income tax rates on people with higher incomes.
 - f. Instituting laws against driving whilst under the influence of alcohol.
- 10 In what ways is your standard of living different from that of your parents or grandparents when they were your age? Why have these changes occurred?

2

THINKING LIKE AN ECONOMIST

INTRODUCTION

Every field of study has its own language, its own processes, its methods of discovery and its own way of thinking. Economics is no different. As you embark on your study of economics, the understanding you bring to the discipline is going to be very different to that which your lecturer has. You will have to learn lots of new terms as is the case in any new subject area. You will also have to learn how economists go about their work and how new ideas in the subject are developed and refined over time. One of the challenges facing students of economics is that many terms used are also used in everyday language. In economics, however, these terms mean specific things. The challenge, therefore, is to set aside that everyday understanding and think of the term or concept as economists do.

Many of the concepts you will come across in this book are abstract. Abstract concepts are ones which are not concrete or real – they have no tangible qualities. We will talk about markets, efficiency, comparative advantage and equilibrium, for example, but it is not easy to physically see these concepts. There are also some concepts that are fundamental to the subject – if you master these concepts they act as a portal which enables you to think like an economist. Once you have mastered these concepts you will never think in the same way again and you will never look at an issue in the same way.

These concepts are referred to as *threshold concepts*. You can read about this further in Meyer et al. (Meyer, J.H.F. and Land, R. (2005). 'Threshold concepts and troublesome knowledge 2: epistemological considerations and a conceptual framework for teaching and learning'. *Higher Education*, 49: 373–388). As you work through your modules you will find that it is not always easy to think like an economist and that there will be times when you are confused, find some of the ideas and concepts being presented to you running contrary to common sense (i.e. they are counter intuitive). What you will be experiencing is what is called *troublesome knowledge*. Don't worry about this – what you are experiencing is perfectly normal and a part of the learning journey. As you travel along this learning journey you will be provided with new information and as a result develop new and useful ways of thinking about the world in which you live.

This chapter discusses the field's methodology. What is distinctive about how economists confront a question? What does it mean to think like an economist? What tools do economists use to explain the world we live in?

THE ECONOMIST AS SCIENTIST

Economists try to address their subject with a scientist's objectivity. They approach the study of the economy in much the same way as a physicist approaches the study of matter and a biologist approaches the study of life: they devise theories, collect data and then analyse these data in an attempt to verify or refute their theories.

There is much debate about whether economics can ever be a science – principally because it is dealing with human behaviour. The essence of any science is *scientific method* – the dispassionate development and testing of theories about how the world works. This method of inquiry is as applicable to studying a nation's economy as it is to studying the Earth's gravity or a species' evolution. As Albert Einstein once put it, 'The whole of science is nothing more than the refinement of everyday thinking'.

Empiricism

One of the *Ten Principles of Economics* states that 'prices rise when the government prints too much money'. To make such a statement we have to have some evidence that this is indeed the case. How did economists find out that printing too much money leads to rising prices? The evidence for this principle is empirical. Empirical means that information has been gathered either by observation, experience or experiment of an event or phenomena (such as a period of rising prices), the formulation of a hypothesis (that prices rise when government prints too much money) and the testing of the hypothesis. A hypothesis is an assumption; the word is derived from the Greek (*hypotithenai*) meaning 'to suppose'. A hypothesis can be developed through observation or experience of phenomena or through what we might call 'idle reasoning'. Having developed a hypothesis, the economist might use scientific method to test it to see whether the hypothesis can be supported, rejected or there might be no evidence to support the hypothesis either way, or they might apply inductive reasoning to explain it.

Inductive and Deductive Reasoning Inductive reasoning refers to the process of observation from which patterns might be formed which provides evidence for a hypothesis which may lead to a theory. In contrast, deductive reasoning begins with a theory from which a hypothesis is drawn. The hypothesis is then subject to observation and either confirmation or rejection. One is not any better than the other – they are different ways of approaching research and may be closely linked. What is important is that we always treat any research with a degree of critical awareness – we don't just simply accept the conclusions drawn from the research but question them and subject them to further testing. Through this circular process, refinements and improvements to theories and explanations can be developed which in turn allow us to make more informed decisions.

A classic example of the relationship between inductive and deductive reasoning can be seen in the case of observing swans. The researcher observes a river with swans swimming past. Every swan observed is white. At the end of the observation period the researcher draws a conclusion that 'all swans are white'. The evidence gathered does support the hypothesis that 'all swans are white'. A theory about why swans are always white might be developed to explain this phenomenon. Subsequent testing of the hypothesis might confirm that, based on the evidence, all swans are white and this might be the accepted hypothesis for many years until, one day, someone sees a black swan. At which point the hypothesis is rejected and the theory will have to be modified. This may lead the researcher to begin asking questions about why the majority of swans appear to be white and what reasons there might be for some swans being black which can again be the subject of empirical research.

Theories Throughout this book we will look at theories. Theories can be used to explain something and to make predictions. The theory of indifference curves and budget lines can be used to explain consumer behaviour. Theories, however, can be developed separate to empirical research. In 1982, Nobel prize-winning economist, Wassily Leontif, lamented the lack of systematic empirical enquiry in economics at the expense of too much 'theorizing'. Leontif (Leontif, W. (1982) *Academic Economics*. In *Science* 217: 104–107) had looked at articles in the *American Economic Review* and observed that a large proportion of the articles contained models which were not supported by any data and analysed issues without supporting data. In comparison, articles based on primary data generated by the author/s, or on secondary data and which used appropriate statistical tests to arrive at conclusions, represented the minority. Was it wise to base decisions or policy on the knowledge generated by research which was not subject to the rigours of empirical methods?

Theorizing on its own can be said to be a tradition in rationalist economics. Here, the economist uses logic, reason and induction to arrive at conclusions. Much of the logic might be based on assumptions which may not be subject to any supporting data. For example, the theory of consumer behaviour makes assumptions about the way human beings behave when making consumption decisions, such as consumers act rationally, prefer more to less and make purchasing decisions based on pure self-interest.

Empiricist economics starts with observations and data from which models can be developed which reflect the data. These models can be used to arrive at conclusions and to make predictions. Observations of consumer behaviour suggest that humans do not act rationally when making purchasing decisions. Models can be developed which represent the data – how humans *do* behave as consumers – and these

models can be used to generate theories which help make predictions about consumer behaviour in different situations.

Even if the hypothesis is supported by the evidence this is not a reason for the economist to sit back and relax, happy in the knowledge they have found ‘the truth’. Things change, new information, experience or observations may be made which render the original hypothesis redundant and subject to revision or refinement; there is a process which is never ending.

The Scientific Method: Observation, Theory and More Observation

The interplay between theory and observation is central to the methodology adopted in the field of economics. Let us return to the principle that prices rise when the government prints too much money. How might this principle have arisen? An economist might live in a country experiencing rapid increases in prices and be moved by this observation to develop a theory of inflation. The theory might assert that high inflation arises when the government prints too much money. To test this theory, the economist could collect and analyse data on prices and money from many different countries. If growth in the quantity of money were not at all related to the rate at which prices are rising, the economist would start to doubt the validity of his theory of inflation. If money growth and inflation were strongly correlated in international data, as in fact they are, the economist would become more confident in his theory.

Although economists use theory and observation like other scientists, they do face an obstacle that makes their task especially challenging: experiments are often difficult in economics. Physicists studying gravity can drop many objects in their laboratories to generate data to test their theories. By contrast, economists studying inflation are not allowed to manipulate a nation’s monetary policy simply to generate useful data. Economists, like astronomers and evolutionary biologists, usually have to make do with whatever data the world happens to give them.

To find a substitute for laboratory experiments, economists pay close attention to the natural experiments offered by history. When a war in the Middle East interrupts the flow of crude oil, for instance, oil prices shoot up around the world. For consumers of oil and oil products, such an event depresses living standards. For economic policymakers, it poses a difficult choice about how best to respond. But for economic scientists, it provides an opportunity to study the effects of a key natural resource on the world’s economies, and this opportunity persists long after the wartime increase in oil prices is over. Throughout this book, therefore, we consider many historical episodes. These episodes are valuable to study because they give us insight into the economy of the past and, more important, because they allow us to illustrate and evaluate economic theories of the present.

Empiricism or Rationalism?

Like many things in economics, there is no right answer to this question. The debate between rationalists is beyond the scope of this introduction. The debate lies in the realm of economic philosophy – about the nature of knowledge, how we acquire knowledge. Some things may seem intuitive and have led to ‘laws’ or beliefs to be widely held and in some cases entrenched. For example, intuition tells us that if the price of a good rises, people will buy less of that good. Rationally, this would also make sense and logic would tell us that if the price of a good is higher it changes the way people view that good, whether they can now afford it and how they view the good in relation to others that they could buy that are similar. Putting all these things together the conclusion must be that when prices rise, the amount purchased will fall. Such a conclusion can then be seen as being generalizable. This means that we can extend the reasoning for this one good to most others.

Some of these can become self-evident truths which are accepted by the general public. For example, the idea that a large proportion of those who claim welfare benefits are ‘scroungers’, that an increase in immigration will take jobs away from the indigenous population, that the increase in carbon emissions will cause global warming, or that the way of reducing the budget deficit is to cut taxes, may all be widely accepted ‘truths’.

Economists will be mindful of the debate between rationalists and empiricists but the idea that scientific methodology can be applied to economics as a discipline and help to improve the knowledge we have and to build upon this knowledge, regardless of how we arrived at it (whether through initial

observation or through theorizing) is now widely accepted as the direction for the subject. The Cambridge economist, Joan Robinson, perhaps captured the debate very well when she wrote [economics] ‘limps along with one foot in untested hypotheses and the other in untestable slogans ... our task is to sort out as best we may this mixture of ideology and science’ (Robinson, J. (1968) *Economic Philosophy*. Pelican).

The Role of Assumptions

If you ask a physicist how long it would take for a cannonball to fall from the top of the Leaning Tower of Pisa, she will probably answer the question by assuming that the cannonball falls in a vacuum. Of course, this assumption is false. In fact, the building is surrounded by air, which exerts friction on the falling cannonball and slows it down. Yet the physicist will correctly point out that friction on the cannonball is so small in relation to its weight that its effect is negligible. Assuming the cannonball falls in a vacuum greatly simplifies the problem without substantially affecting the answer.

Economists make assumptions for the same reason: assumptions can simplify the complex world and make it easier to understand. To study the effects of international trade, for example, we may assume that the world consists of only two countries and that each country produces only two goods. Of course, the real world consists of dozens of countries, each of which produces thousands of different types of goods. But by assuming two countries and two goods, we can focus our thinking. Once we understand international trade in an imaginary world with two countries and two goods, we are in a better position to understand international trade in the more complex world in which we live.

The art in scientific thinking is deciding which assumptions to make. Suppose, for instance, that we were dropping a beach ball rather than a cannonball from the top of the building. Our physicist would realize that the assumption of no friction is far less accurate in this case: friction exerts a greater force on a beach ball than on a cannonball because a beach ball is much larger and, moreover, the effects of air friction may not be negligible relative to the weight of the ball because it is so light. The assumption that gravity works in a vacuum may, therefore, be reasonable for studying a falling cannonball but not for studying a falling beach ball.

Similarly, economists use different assumptions to answer different questions. What happens to demand if price changes but we assume incomes stay constant? If we make an assumption of perfect competition in an industry, what will happen to long-run profits? What will happen to the economy when the government changes the amount of money in circulation if we assume prices do not change much in the short run? How does this differ in the long run if we assume that prices are flexible? As we have seen above, these assumptions have to be used with care and they are also subject to testing to see the extent to which the assumption is reasonable in the same way that it is deemed reasonable by the physicist to drop the assumption of friction when considering the effect of dropping a cannonball from the Leaning Tower of Pisa.

Experiments in Economics

Because economics is a science which is centred on human behaviour, it is not always possible to conduct experiments in the way in which physical sciences like Biology, Chemistry and Physics do. However, there are two major fields of experimentation in economics that are worthy of note. Experiments in economics can be conducted in a ‘laboratory’ where data can be collected via observations on individual or group behaviour, through questionnaires and surveys, interviews and so on, or through the collection and analysis of data that exists such as wages, prices, stock prices and volumes of trades, unemployment levels, inflation and so on. The data can be analysed in relation to a research question and conclusions drawn which help develop new understanding or refine and improve existing understanding. The conclusions drawn from such experiments may be generalizable; in other words, the findings of the experiment can be extended outside the ‘laboratory’ to explain behaviour or economic phenomena and provide the basis for prediction.

One example of how such laboratory experiments can help change understanding is the work of people such as Daniel Kahneman, Amos Tversky, Richard Thaler and Cass Sunstein, whose research has helped to provide insights into judgement and decision making and has offered a different perspective on the assumptions of rational decision making. Thaler, for example, conducted a number of experiments to explore how individuals respond when faced with different questions on losses and gains in relation to a reference point. He found that prior ownership of a good, for example a ticket to see a football game, altered people’s willingness to sell, even at prices significantly higher than that which they had paid. Thaler’s

observations on the consistency of this behaviour across a number of experiments led him to coin the term *endowment effect* to explain the behaviour and it is now widely accepted that the endowment effect does exist and that it runs counter to the assumptions of rational behaviour in economics. We will look at this in more detail in Chapter 5. Thaler then worked with Kahneman and Tversky and extended the theory to distinguish between goods which are held for trade and those which are held for use. The endowment effect, they suggested, was not universal, it was more powerful when goods were held for use.

A second type of experiment in economics is *natural experiments*. A natural experiment is one where the study of phenomena is determined by natural conditions which are not in the control of the experimenter. Natural experiments can be exploited when some change occurs which allows observation to be carried out on the effects of this change in one population and comparisons made with another population who are not affected. Examples of natural experiments include observing the effects of bans on smoking in public places on the number of people smoking or the possible health benefits, how far a change in the way in which education is financed affects standards, the effect of the length of schooling and income, the effects of a rise in a tax on property on the market for housing, what are the effects on the female labour market of changes in fertility treatment and availability?

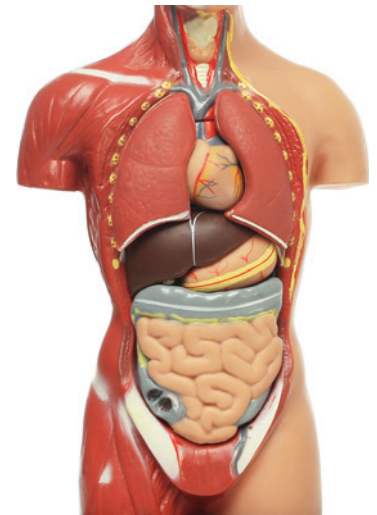
Typically, natural experiments make use of the statistical tools of correlation and regression to determine whether there is any relationship between two or more variables, what the strength of the relationship might be, if any such relationship exists and what the nature of the relationship is. From such analysis, a model can be developed which can be used to predict. At the heart of such analysis is the extent to which a relationship between two or more variables can be linked to cause and effect. Just because two variables appear to have some relationship does not necessarily imply cause and effect. For example, a researcher might find that an observation of graduates in the workforce shows that their incomes are generally higher than those of non-graduates; can the researcher conclude that having a degree will lead to higher income? Possibly, but not necessarily. There might be other factors that have an effect on income apart from having a degree and trying to build a model which takes into account these different factors is an important part of the value of natural experiments.

Models in Economics

The image on the right shows a plastic replica of the human body. Anatomical models like the one shown have all the major organs – the heart, liver, kidneys and so on – and are used in teaching anatomy and biology. The models allow teachers to show students in a simple way how the important parts of the body fit together. Of course, these plastic models are not actual human bodies; they are stylized and omit many details. Despite this, using these models is useful for learning how the human body works. Economists also use models to learn about the world, but, instead of being made of plastic, they are most often composed of diagrams and equations. It is important to remember that we should not confuse economic models with reality just as no-one would mistake an anatomical model for a real person. Economic models omit many details to allow us to see what is truly important. As we use models to examine various economic issues throughout this book, you will see that all the models are built with assumptions. In making these assumptions we can focus on the specific things that we want to study. For example, if we are studying a model of the market we might make an assumption that supply remains constant and that the factors affecting demand apart from a change in income are also constant. By simplifying reality in this way through the model, we can improve our understanding of it.

A model will contain a number of variables. In the example above the variables in the market model are demand, supply, incomes, tastes, price and so on. Some of these variables are determined by the model and some are generated within the model. For example, take the market model where the quantity demanded (Q_d) is dependent on the price. Q_d is said to be the dependent variable. Its value, however, will be dependent on the functional

Models are used in many different disciplines. This anatomical model is used to help students understand the workings of the human body and the major organs. In Economics, models are similar stylized representations of reality.



relationships in the model (the factors that affect demand) such as incomes, tastes and the prices of other goods. Q_d can be described as an **endogenous variable**. Price, on the other hand, is the independent variable – it affects the model (the quantity demanded) but is not affected by it. The price is not determined by, or dependent on, the quantity demanded. Price would be referred to as an **exogenous variable**.

endogenous variable a variable whose value is determined within the model
exogenous variable a variable whose value is determined outside the model

Understanding the difference between endogenous and exogenous variables is important because it helps us to separate out cause and effect. Does a change in price, for example, cause a change in quantity demanded or does it affect quantity demanded?

Our First Model: The Circular-Flow Diagram

The economy consists of millions of people engaged in many activities – buying, selling, working, hiring, manufacturing and so on. It is extremely complex. It is extremely difficult to intuitively understand how the economy works. A model helps to simplify our thinking about all these activities.

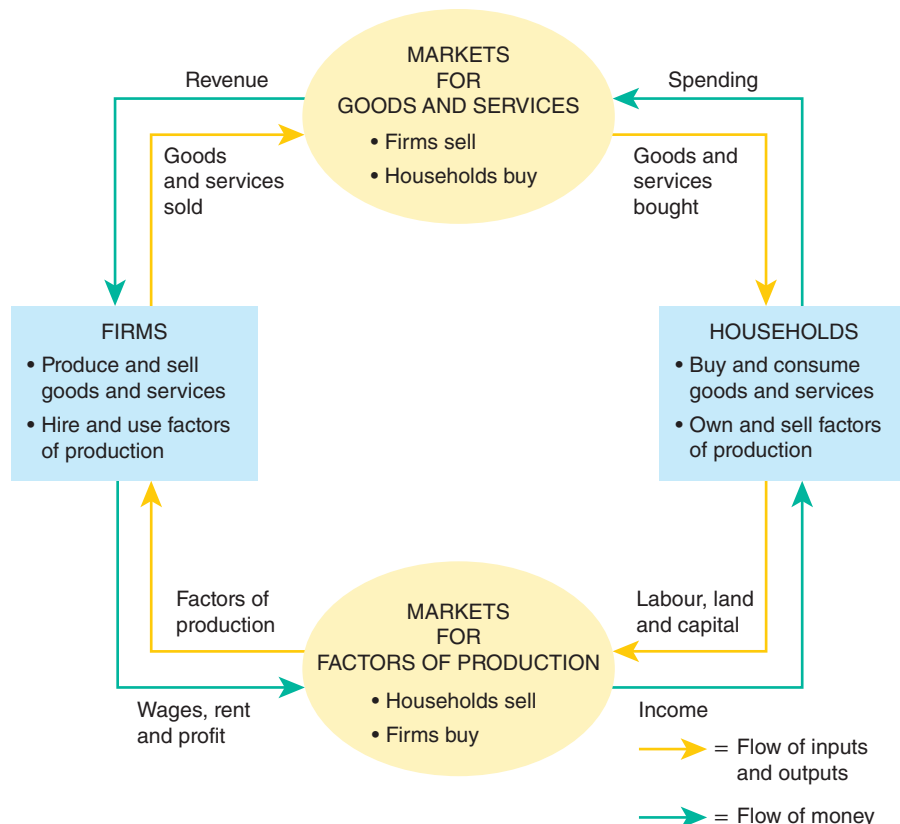
Figure 2.1 presents a visual model of the economy, called a **circular-flow diagram**. In this model, the economy is simplified to include only two types of decision makers – households and firms. Firms produce goods and services using the factor inputs of labour, land and capital. Households own the factors of production and consume all the goods and services that the firms produce.

circular-flow diagram a visual model of the economy that shows how money and production inputs and outputs flow through markets among households and firms

FIGURE 2.1

The Circular Flow

This diagram is a schematic representation or model of the organization of the economy. Decisions are made by households and firms. Households and firms interact in the markets for goods and services (where households are buyers and firms are sellers) and in the markets for the factors of production (where firms are buyers and households are sellers). The outer set of arrows shows the flow of money, and the inner set of arrows shows the corresponding flow of inputs and outputs.



Households and firms interact in two types of markets. In the *markets for goods and services*, households are buyers and firms are sellers. In particular, households buy the output of goods and services that firms produce. In the *markets for the factors of production*, households are sellers and firms are buyers. In these markets, households provide the inputs that the firms use to produce goods and services. The circular-flow diagram offers a simple way of representing the organization of all the economic transactions that occurs between households and firms in the economy.

The inner loop of the circular-flow diagram represents the flows of inputs and outputs. The households sell the use of their labour, land and capital to the firms in the markets for the factors of production. The firms then use these factors to produce goods and services, which in turn are sold to households in the markets for goods and services. Hence, the factors of production flow from households to firms, and goods and services flow from firms to households.

The outer loop of the circular-flow diagram represents the corresponding flows of money. The households spend money to buy goods and services from the firms. The firms use some of the revenue from these sales to pay for the factors of production, such as the wages of their workers. What's left is the profit of the firm's owners, who themselves are members of households. Hence, spending on goods and services flows from households to firms, and income, in the form of wages, rent and profit, flows from firms to households.

Let's take a tour of the circular flow by following a one euro coin as it makes its way from person to person through the economy. Imagine that the euro begins at a household, sitting in, say, your pocket. If you want to buy a cup of coffee, you take the euro to one of the economy's markets for goods and services, such as your local café. There you spend it on your favourite drink: a double espresso. When the euro moves into the café cash register, it becomes revenue for the owner of the café. The euro doesn't stay with the café owner for long, however, because he uses it to buy inputs in the markets for the factors of production. For instance, the café owner might use the euro to pay rent to the owner of the building that the café occupies or to pay the wages of its workers. In either case, the euro enters the income of some household and, once again, is back in someone's pocket. At that point, the story of the economy's circular flow starts once again.

The circular-flow diagram in Figure 2.1 is one simple model of the economy. It is useful for developing some basic ideas about how the economy works. It dispenses with details that, for some purposes, might be significant. We will introduce a more complex circular flow model later in the book which takes account of taxes, government spending, saving, investment and exports and imports.

THE ECONOMIST AS POLICY ADVISOR

Often economists are asked to explain the causes of economic events. Why, for example, is unemployment higher for teenagers than for older workers? Sometimes economists are asked to recommend policies to improve economic outcomes. What, for instance, should the government do to improve the economic well-being of teenagers? When economists are trying to explain the world, they are scientists. When they are trying to help improve it, they are policy advisors.

Positive versus Normative Analysis

To help clarify the two roles that economists play, we begin by examining the use of language. Because scientists and policy advisors have different goals, they use language in different ways.

For example, suppose that two people are discussing minimum wage laws.

Pascale: Minimum wage laws cause unemployment

Sophie: The government should raise the minimum wage

There is a fundamental difference in these two statements. Pascale's statement is spoken like that of a scientist: she is making a claim about how the world works. Sophie is speaking like a policy advisor: she is making a claim about how she would like to change the world.

Pascale's is making a positive statement. **Positive statements** are descriptive. They make a claim about how the world *is*. Positive statements have the property that the claims in them can be tested and confirmed, refuted or shown to not be provable either way. A second type of statement, such as Sophie's,

is normative. **Normative statements** are prescriptive. They make a claim about how the world *ought to be*. Normative statements have the property that they include opinion; it is not possible to test opinions and confirm or reject them.

Positive statements claims that attempt to describe the world as it is
Normative statements claims that attempt to prescribe how the world should be

In studying economics (and in everyday life) you will come across many examples of positive and normative statements. In conducting analysis it is important to distinguish between the two. It is perfectly possible to conduct both positive and normative analysis. For example, the statement: *the government should reduce the deficit as this will benefit the economy*, is a normative statement – it contains an opinion that the government ought to reduce the deficit. *A reduction in the government deficit will benefit the economy* is a positive statement, it is capable of being tested. Economists could engage in positive analysis to test whether there is any evidence to support the statement but equally could engage in normative analysis on the basis that there are many people who believe that reducing the deficit will benefit the economy.

A key difference between positive and normative statements, therefore, is how we judge their validity. An economist might evaluate Pascale's statement by analysing data on changes in minimum wages and changes in unemployment over time. By contrast, evaluating normative statements involves values as well as facts. Sophie's statement cannot be judged using data alone. Deciding what is good or bad policy is not merely a matter of science; it also involves our views on ethics, religion and political philosophy.

Of course, positive and normative statements may be related. Our positive views about how the world works affect our normative views about what policies are desirable. Pascale's claim that the minimum wage causes unemployment, if true, might lead us to reject Sophie's conclusion that the government should raise the minimum wage. Yet our normative conclusions cannot come from positive analysis alone; they involve value judgements as well. Normative analysis has its value but it may be necessary to carry out positive analysis first in order to inform the normative.

As you study economics, keep in mind the distinction between positive and normative statements. Much of economics just tries to explain how the economy works. Yet often the goal of economics is to improve how the economy works. When you hear economists making normative statements, you know they have crossed the line from scientist to policy advisor.

WHY ECONOMISTS DISAGREE

There are many jokes about economists and in recent years the profession has come under scrutiny in the wake of the financial crisis. Some of these jokes imply that economists' advice is either contradictory or not definite. One joke about economists concerns the politician who announced that in future she would only employ economists with one hand. The reason being she was tired of receiving advice from her economists which said: 'On the one hand ... On the other hand ...'

A witticism from the Irish playwright and co-founder of the London School of Economics goes: 'If all economists were laid end to end, they would not reach a conclusion'. Economists as a group are often criticized for giving conflicting advice to policymakers. There are two basic reasons:

- Economists may disagree about the validity of alternative positive theories about how the world works.
- Economists may have different values and, therefore, different normative views about what policy should try to accomplish.

Let's discuss each of these reasons.

Differences in Scientific Judgements

Several centuries ago, astronomers debated whether the Earth or the Sun was at the centre of the solar system. More recently, meteorologists have debated whether the Earth is experiencing global warming

and, if so, why. Science is a search for understanding about the world around us. It is not surprising that as the search continues, scientists can disagree about the direction in which truth lies.

Economists often disagree for the same reason. Economics is a young science, and there is still much to be learned. Economists sometimes disagree because they have different beliefs about the validity of alternative theories or about the size of important parameters.

For example, economists disagree about whether the government should levy taxes based on a household's income or based on its consumption (spending). Advocates of a switch from an income tax to a consumption tax believe that the change would encourage households to save more, because income that is saved would not be taxed. Higher saving, in turn, would lead to more rapid growth in productivity and living standards. Advocates of an income tax system believe that household saving would not respond much to a change in the tax laws. These two groups of economists hold different normative views about the tax system because they have different positive views about the responsiveness of saving to tax incentives.

Differences in Values

Anneka and Henrik both take water from the town well. To pay for maintaining the well, the town imposes a property tax on its residents. Anneka lives in a large house worth €2 million and pays a property tax of €10,000 a year. Henrik owns a small cottage worth €20,000 and pays a property tax of €1,000 a year.

Is this policy fair? If not, who pays too much and who pays too little? Would it be better to replace the tax based on the value of the property with a tax that was just a single payment from everyone living in the town (a poll tax) in return for using the well – say, €1,000 a year? After all, Anneka lives on her own and actually uses much less water than Henrik and the other four members of his family who live with him and use more water as a result. Would that be a fairer policy?

This raises two interesting questions in economics – how do we define words like 'fair' and 'unfair', and who holds the power to influence and make decisions? If the power is in the hands of certain groups in government or powerful businesses, the policies may be adopted even if they are widely perceived as being 'unfair'.

What about replacing the property tax not with a poll tax but with an income tax? Anneka has an income of €100,000 a year so that a 5 per cent income tax would present her with a tax bill of €5,000. Henrik, on the other hand, has an income of only €10,000 a year and so would pay only €500 a year in tax and the members of his family who do not work don't pay any income tax. Does it matter whether Henrik's low income is due to his decision not to go to university and take a low paid job? Would it matter if it were due to a physical disability? Does it matter whether Anneka's high income is due to a large inheritance from her family? What if it were due to her willingness to work long hours at a dreary job?

These are difficult questions on which people are likely to disagree. If the town hired two experts to study how the town should tax its residents to pay for the well, we should not be surprised if they offered conflicting advice.

This simple example shows why economists sometimes disagree about public policy. As we learned earlier in our discussion of normative and positive analysis, policies cannot be judged on scientific grounds alone. Economists give conflicting advice sometimes because they have different values. Perfecting the science of economics will not tell us whether it is Anneka or Henrik who pays too much.

Perception versus Reality

Because of differences in scientific judgements and differences in values, some disagreement among economists is inevitable. Yet one should not overstate the amount of disagreement. In many cases, economists do offer a united view.

Table 2.1 contains ten propositions about economic policy. In a survey of economists in business, government and academia, these propositions were endorsed by an overwhelming majority of respondents. Most of these propositions would fail to command a similar consensus among the general public.

One of the propositions in the table concerns tariffs and import quotas, two policies that restrict trade among nations. For reasons we will discuss more fully in later chapters, almost all economists oppose

such barriers to free trade. This, in fact, is one of the major reasons why the European Union (EU) was set up and why countries are queuing up to join it: tariffs and quotas are not imposed on trade between EU member countries. Tariffs and quotas *are*, however, often imposed by EU countries on goods coming from outside of the European Union.

TABLE 2.1**Ten Propositions About Which Most Economists Agree***Proposition (and percentage of economists who agree)*

-
1. A ceiling on rents reduces the quantity and quality of housing available. (93%)
 2. Tariffs and import quotas usually reduce general economic welfare. (93%)
 3. Flexible and floating exchange rates offer an effective international monetary arrangement. (90%)
 4. Fiscal policy (e.g., tax cut and/or government expenditure increase) has a significant stimulative impact on a less than fully employed economy. (90%)
 5. If the government budget is to be balanced, it should be done over the business cycle rather than yearly. (85%)
 6. Cash payments increase the welfare of recipients to a greater degree than do transfers-in-kind of equal cash value. (84%)
 7. A large government budget deficit has an adverse effect on the economy. (83%)
 8. A minimum wage increases unemployment among young and unskilled workers. (79%)
 9. The government should restructure the welfare system along the lines of a 'negative income tax'. (79%)
 10. Effluent taxes and marketable pollution permits represent a better approach to pollution control than imposition of pollution ceilings. (78%)
-

Another of the propositions concerns the imposition of a legal minimum wage – nearly 80 per cent of economists surveyed said that they thought a minimum wage increases unemployment among unskilled and young workers. Nevertheless, the majority of European Union countries now have a statutory minimum wage. Of course, these economists were not necessarily *against* the imposition of a minimum wage. Some of them might argue, for example, that while, on the one hand, introducing a minimum wage above a certain level may affect unemployment, on the other hand it may increase the average quality of goods and services produced in the economy by making it harder for producers of low-quality goods and services to compete by keeping wages and prices low, and this may lead to a net benefit to the economy overall. Remember: people face trade-offs.

Table 2.2 shows another aspect of this same idea. This is based on a paper written by Alan Budd, a former Chief Economic Advisor to HM Treasury in the UK, former Chair of the Office for Budget Responsibility and Provost of the Queen's College, Oxford University until 2008. The paper appeared in the journal *World Economics* in 2004. This table shows a list of things that Sir Alan thinks that economists know and therefore agree on. Most of these things are based on hypotheses that have been tested and observed and which appear to explain key behaviours. It is worth bearing the information in Tables 2.1 and 2.2 in mind as you read through the book.

TABLE 2.2**What Economists Know****Economists know that demand curves slope downwards from left to right.**

- This suggests that when prices rise, the quantity demanded of a product (under normal circumstances) will fall and vice versa.

Economists know that supply curves slope upwards from left to right.

- This suggests that when prices rise, the quantity supplied by producers will also rise and vice versa.

Economists know that the proportion of income spent on food falls as income rises.

- This is what is referred to as Engel's law, after the German statistician, Ernst Engel. Budd points out that this law applies not only to studies of the differences in the proportion of income spent on food by the rich and the poor, but also over time. When nations go through economic development and become richer, the proportion of national income spent on food falls.

Economists know that there are gains from trade to be had when countries or individuals have comparative advantage.

- What this means is that one country might be better at producing a number of goods compared to another. However, if it focuses its attention on producing the good in which it has a comparative advantage, both countries can gain from trade and the world economy will be in a better state. By focusing on production of one good at the expense of others, countries move resources from the production of one good to another. By doing so, the country sacrifices the output of this other good that it could have produced. However, the gains made from shifting resources into production of products in which they are more efficient in production helps to raise total output. A mutually beneficial rate of exchange between the two countries means they both are better off than before.

Economists tend to think in terms of general rather than partial equilibrium.

- Economies are made up of millions of interrelated markets. Non-economists, argues Budd, tend to see things from a partial equilibrium point of view. In many cases, this view is based around a zero-sum outcome – the benefit received by one party to an economic decision is offset by a negative impact on someone else. Looking at the big picture gives a more accurate understanding of how economies work and what the consequences of economic policy can be.

Economists know that sunk costs should not affect pricing decisions.

- Sunk costs are those costs that have been spent and cannot be recovered or affected by future economic activity. What this means is that the spending on (for example) major infrastructure projects will be largely academic. Arguments about how much it will cost to build a new high-speed rail link for the eastern half of the UK are simply not relevant. What will be important are the on-going costs of actually running the line and the infrastructure that is built.

There is a difference between what economists know and what non-economists know.

- This can be summarized under the heading 'folk economics'. Folk economics is the intuitive understanding that untrained people have about how the economy operates. If you like, this can be seen as being a novice perspective as opposed to an expert one.

Budd makes an interesting observation here. If you (as a non-physicist) were engaged in conversation with a physicist who tells you about string theory, you are unlikely to interrupt him or her and disagree with what they were telling you. An economist engaged in a discussion about obesity, for example, with a non-economist would be far more likely to be questioned about the views being put forward. If I, as an economist, told you, a non-economist, that obesity was caused by the rise in poverty levels in countries, it is likely that there would be some disagreement with my view. In short, economists think in the subject, whereas non-economists do not!

SELF TEST Why might economic advisors to the government disagree about a question of policy such as reducing a budget deficit?

Economists as Decision Makers

It could be said that economics is the science of decision making. The way that economists go about making decisions is a specific one. They will initially try to identify the problem or issue related to the decision: for example, will measures to cut greenhouse gas emissions be efficient; or is it worth my while travelling 50 miles to redeem a voucher for €50 at a particular store?

The next stage is to look at the costs and benefits involved in the decision. These costs and benefits are not just the private costs and benefits to the individual concerned; however, they will also include the costs and benefits to third parties who are not directly involved in the actual decision. For example, cutting greenhouse gas emission means that resources will have to be diverted to new ways of production or different ways of producing energy. The private costs will be those borne by the businesses that will have to implement measures to adhere to the limits placed upon them. The social costs might include the impact on local people of the construction of wind farms or new nuclear power stations. If I chose to make the 50-mile journey then I would incur travelling costs – petrol, vehicle depreciation and so on – as well as the cost in terms of the time I have to give to make the journey. The social costs include the addition to possible road congestion that I add, as well as the potential danger to other road users that my being on the road presents.

Having identified the costs and benefits, the economist then seeks to place a value on them in order to get some idea of the relationship between the costs and benefits of making the decision. In some cases, valuing costs and benefits can be easy: in the case of my shopping trip, the benefit is the €50 saved; the costs of petrol used on the journey are easy to calculate. Some costs and benefits are much more difficult to value. The loss of visual amenity for a resident living near a wind turbine or the value of the possible loss of life from a nuclear catastrophe at a power plant may be very difficult to value. Economists have devised ways in which these values can be estimated.

Once the sum of the costs and benefits are calculated, the decision then becomes clearer. If the cost outweighs the benefit then making the decision may be unwise, but if the costs are less than the benefits then it may mean the decision is warranted. Policymakers may want to look at the extent to which the costs outweigh the benefit or the benefit outweighs the costs, however. If the benefit to me of going to redeem the voucher is €50 but the cost of making the trip to do so is valued at €49 then it may not be worth my while going, but if the cost were valued at just €10 then my decision may be more obvious.

Every day millions of decisions are made by individuals, businesses and governments. Whilst not every one of these decisions will be made using the exact processes outlined above, and we certainly do not stop to think about how we rationalize our decisions, nevertheless our brains do engage in computational processes as we make decisions, but they are mostly subconscious. Economists and psychologists are increasingly finding out more about how humans make decisions, which is helping improve our understanding of the models which we use to analyse consumer behaviour.

LET'S GET GOING

The first two chapters of this book have introduced you to the ideas and methods of economics. We are now ready to get to work. In the next chapter we start learning in more detail the principles of economic behaviour and economic policy.

As you proceed through this book, you will be asked to draw on many of your intellectual skills. You might find it helpful to keep in mind some advice from the great John Maynard Keynes:

The study of economics does not seem to require any specialised gifts of an unusually high order. Is it not ... a very easy subject compared with the higher branches of philosophy or pure science? An easy subject, at which very few excel! The paradox finds its explanation, perhaps, in that the master-economist must possess a rare combination of gifts. He must be mathematician, historian, statesman, philosopher – in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purposes of the future. No part of man's nature or his institutions must lie entirely outside his regard. He must be purposeful and disinterested

in a simultaneous mood; as aloof and incorruptible as an artist, yet sometimes as near the earth as a politician.

It is a tall order. But with practice you will become more and more accustomed to thinking like an economist.

SUMMARY

- Economists try to address their subject with a scientist's objectivity. Like all scientists, they make appropriate assumptions and build simplified models in order to understand the world around them. One simple economic model is the circular-flow diagram.
- Economists use empirical methods to develop and test hypotheses
- Research can be conducted through using inductive and deductive reasoning – no one way is the 'right way'.
- Economists develop theories which can be used to explain phenomena and make predictions. In developing theories and models, economists have to make assumptions.
- Using theory and observation is part of scientific method but economists always have to remember that they are studying human beings and humans do not always behave in consistent or rational ways.
- A positive statement is an assertion about how the world *is*. A normative statement is an assertion about how the world *ought to be*. When economists make normative statements, they are acting more as policy advisors than scientists.
- Economists who advise policymakers offer conflicting advice either because of differences in scientific judgments or because of differences in values. At other times, economists are united in the advice they offer, but policymakers may choose to ignore it.

QUESTIONS FOR REVIEW

- 1 How is economics like a science?
- 2 Why do economists make assumptions?
- 3 Should an economic model describe reality exactly?
- 4 What is meant by empirical study in economics?
- 5 Using an example, explain the difference between inductive and deductive reasoning.
- 6 Should economic theories be developed as a result of observation or before observation? Explain.
- 7 What is the difference between a positive and a normative statement? Give an example of each.
- 8 What is the role of assumptions in economics?
- 9 Using an example, explain the difference between an endogenous and an exogenous variable.
- 10 Why do economists sometimes offer conflicting advice to policymakers?

PROBLEMS AND APPLICATIONS

- 1 Terms like Investment, Capital, Interest, Price and Cost have different meanings in economics than they do in normal everyday usage. Find out what the differences are and explain why economists might have developed these different meanings.
- 2 One common assumption in economics is that the products of different firms in the same industry are indistinguishable. For each of the following industries, discuss whether this is a reasonable assumption.
 - a. steel
 - b. novels
 - c. wheat
 - d. fast food
 - e. mobile phones (think carefully about this one)
 - f. hairdressers.
- 3 Draw a circular-flow diagram. Identify the parts of the model that correspond to the flow of goods and services and the flow of euros for each of the following activities.
 - a. Sam pays a petrol station €60 for filling up his car with petrol.
 - b. Georgia earns €5.50 per hour working at a fast food restaurant.
 - c. Millie spends €7 to see a film.
 - d. Patrick earns €10,000 from his 10 per cent ownership of Pan-European Industrial.

- 4 A researcher in a university notices that the price of flights to holiday destinations tends to be much higher outside semester dates. He formulates a theory to explain this phenomenon. Has the researcher arrived at the theory by induction or deduction? What steps might the researcher take to apply scientific method to test his theory?
- 5 A politician makes a speech in which he criticizes the government's immigration policy, saying that it is too loose and encourages too many people to enter the country and take jobs away from local people. How might an economist go about assessing the validity of the politician's comments?
- 6 If models are not capable of representing the real world in any detail and rely too much on assumptions, then what value can they be? In the wake of the financial crisis, there has been much criticism of economists' models. Does this suggest that economists need to rethink the way they go about seeking to understand and represent the world?
- 7 Classify each of the following statements as positive or normative. Explain.
 - a. Society faces a short-run trade-off between inflation and unemployment.
 - b. A reduction in the rate of growth of money will reduce the rate of inflation.
 - c. The European Central Bank should reduce the rate of growth of money.
 - d. Society ought to require welfare recipients to look for jobs.
 - e. Lower tax rates encourage more work and more saving.
- 8 Classify each of the statements in Table 2.1 as positive, normative or ambiguous. Explain.
- 9 If you were prime minister, would you be more interested in your economic advisors' positive views or their normative views? Why?
- 10 Would you expect economists to disagree less about public policy as time goes on? Why or why not? Can their differences be completely eliminated? Why or why not?

APPENDIX

Graphing and the Tools of Economics: A Brief Review

As you work through this book and in your course, you will come across a number of standard economic methods and processes which economists use in analysing the world. In this appendix we are going to briefly introduce some of these.

Many of the concepts that economists study can be expressed with numbers – the price of bananas, the quantity of bananas sold, the cost of growing bananas and so on. These numbers often represent variables – things that can change. Economists are interested in these variables and in particular are interested in how variables are related to each other. When the price of bananas rises, for example, people buy fewer bananas. Does this mean that there is a relationship between the variable, price and the variable, demand? If there is sufficient proof that not only are the two variables related but that there is a strong relationship between the two then this may help in predicting a more general relationship between two variables. Economists will use maths to represent such relationships and also through graphs.

Functions

In economics a lot of use is made of functions. Demand and supply equations are two examples of functions. Typically, functions are expressed as:

$$y = f(x)$$

or simply $f(x)$.

This means that the value of y is dependent on the value of the terms in the bracket – in our example above there is only one value, x , so the value of y is dependent on the value of x .

In the example above, the value of y is dependent on the value of a single variable x . However, it is quite possible that the value of y could be dependent on a range of different variables. This can also be represented in the form of a function which would look like:

$$y = f(x_1, \dots, \dots, x_n)$$

where x_1, \dots, \dots, x_n represents a range of variables.

Linear Equations During your course it is likely that you will have to work with linear equations – equations which are represented graphically as straight lines. A linear equation typically looks like:

$$y = a + bx$$

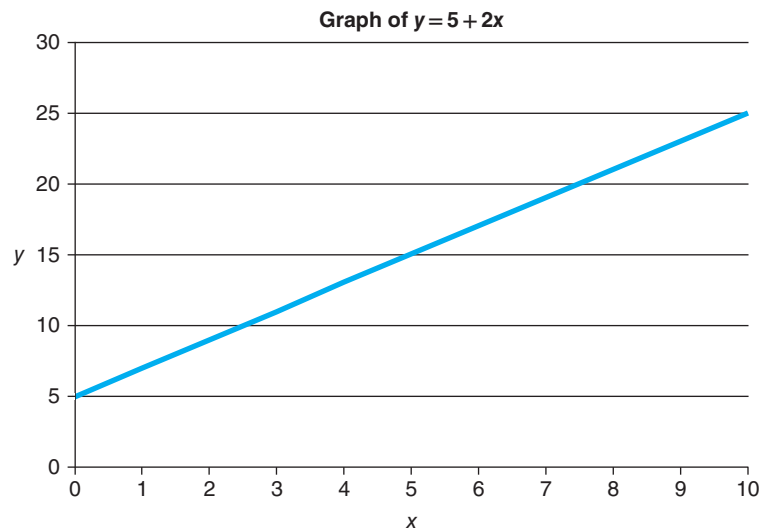
In this equation, y = the value plotted on the vertical axis (the dependent variable) and x is the value on the horizontal axis (the independent variable).

a is a constant and represents the point where the line cuts the y axis and b is the slope of the line or its gradient. We can plot a linear equation as a graph by assigning different values to x and using the equation to establish the value of y in each case. This is represented in the table for Figure 2A.1 for the linear equation $y = 5 + 2x$.

FIGURE 2A.1

Values of x and y for the linear equation $y = 5 + 2x$ between $x = 0$ and $x = 10$ when

| x | $y = 5 + 2x$ | y |
|-----|-----------------|-----|
| 0 | $y = 5 + 2(0)$ | 5 |
| 1 | $y = 5 + 2(1)$ | 7 |
| 2 | $y = 5 + 2(2)$ | 9 |
| 3 | $y = 5 + 2(3)$ | 11 |
| 4 | $y = 5 + 2(4)$ | 13 |
| 5 | $y = 5 + 2(5)$ | 15 |
| 6 | $y = 5 + 2(6)$ | 17 |
| 7 | $y = 5 + 2(7)$ | 19 |
| 8 | $y = 5 + 2(8)$ | 21 |
| 9 | $y = 5 + 2(9)$ | 23 |
| 10 | $y = 5 + 2(10)$ | 25 |



Notice that the line intersects the vertical axis where $x = 0$. The value of the constant a in this equation is 5 and represents the vertical intercept, the point where the equation of the line cuts the vertical axis. Next thing to notice is that as we give different values for x we see that the value of y rises by 2 for each increase of x by 1. In the equation, the constant b is 2 and is the slope or gradient of the line. More on slopes later in this appendix.

Types of Graphs The line graph we generated from the equation is one example of a graph that you will use in your study of economics. There are a number of others.

Why use graphs at all? Graphs serve two purposes. First, when developing economic theories, graphs offer a way to express visually ideas that might be less clear if described with equations or words. Secondly, when analysing economic data, graphs provide a way of finding how variables are in fact related in the world. Whether we are working with theory or with data, graphs provide a means by which we can see patterns and relationships. Choosing the appropriate graphical method is important – the aim is to make the information we are trying to view as clear as possible. An effective economist chooses the type of graph that best suits the purpose at hand.

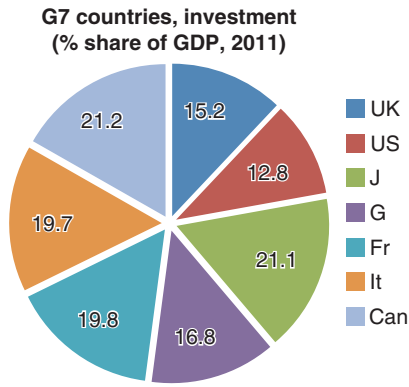
Graphs of a Single Variable

Three common graphs are shown in Figure 2A.2. The *pie chart* in panel (a) shows the proportion of gross domestic product devoted to investment in the so called 'G7 countries'. These countries are Japan, the United States, United Kingdom, France, Germany, Canada and Italy. A slice of the pie represents

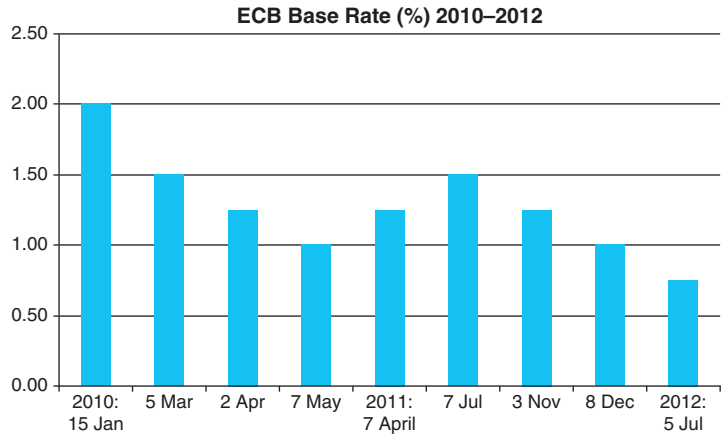
FIGURE 2A.2

Types of Graph

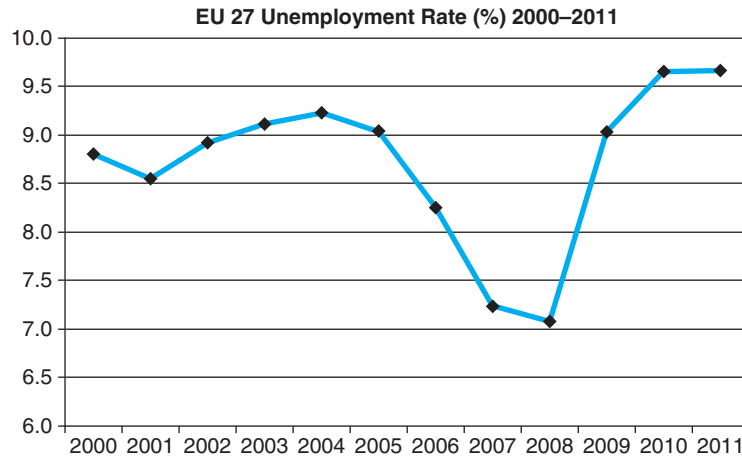
The pie chart in panel (a) shows the proportion of GDP devoted to investment in the G7 countries in 2011. The bar graph in panel (b) shows the interest rate set by the ECB. The time-series graph in panel (c) shows the unemployment rate for the 27 EU countries from 2000–2011. Panel (d) shows the same information as panel (c) but uses a different scale for the vertical axis.



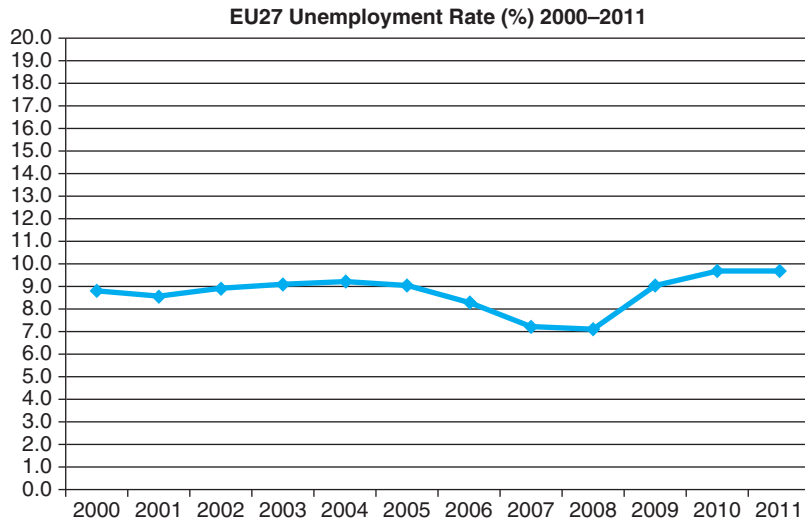
(a)



(b)



(c)



(d)

each country's proportion of investment. The *bar graph* in panel (b) shows the interest rate set by the European Central Bank (ECB) between January 2010 and July 2012. The interest rate is on the vertical axis represented as a percentage and the time period is on the horizontal axis. The height of each bar represents the interest rate. The *time-series graph* in panel (c) traces the unemployment rate in the 27 EU countries between 2000 and 2011. The unemployment rate as a percentage is represented on the vertical axis and the year is on the horizontal axis. The height of the line shows the unemployment rate each year.

Interpreting Line Graphs Look at the line graph in panel (c) of Figure 2A.2. The unemployment rate starts off at about 8.7 per cent in 2000, falls to around 8.5 per cent in 2001 before rising to about 9.2 per cent in 2004. The difference between the unemployment rate in 2001 and 2004 is about 0.7 per cent – a relatively small rise – which is represented by a moderately upward sloping line. From 2004, the unemployment rate falls to around 7.1 per cent in 2008 – a relatively large fall at 2.1 per cent. This is represented by a relatively steep line. Equally, the rise in the unemployment rate after 2008 to around 9.6 per cent in 2010 is represented by a steep rise in the line. Unemployment rose by around 2.5 per cent in just two years compared to the rise in unemployment of just 0.7 per cent in three years between 2001 and 2004. Between 2010 and 2011, the unemployment rate remained at about 9.6 per cent and so we would say that the rate had been 'flat' between those two dates represented by the flat shape of the line.

Looking at the shallowness or steepness of a line graph can tell us a great deal about the rate of change – whether our variable is changing quickly or slowly. However, we must take care in ensuring that we take note of the scale of the graph we are using. Look at the graph in panel (d). This shows exactly the same information as that in panel (c) but the scale on the vertical axis has been changed. The change in the unemployment rate does not look anything like as dramatic in panel (d) and the line appears to be relatively flat. This is a reminder that if we are comparing information on two different line graphs we have to make sure that we take note of the scale before drawing any conclusions.

Graphs of Two Variables: The Coordinate System

Although the three graphs in Figure 2A.2 are useful in showing how a variable changes over time or across individuals, such graphs are limited in how much they can tell us. These graphs display information only on a single variable. Economists are often concerned with the relationships between variables. Thus, they need to be able to display two variables on a single graph. The *coordinate system* makes this possible.

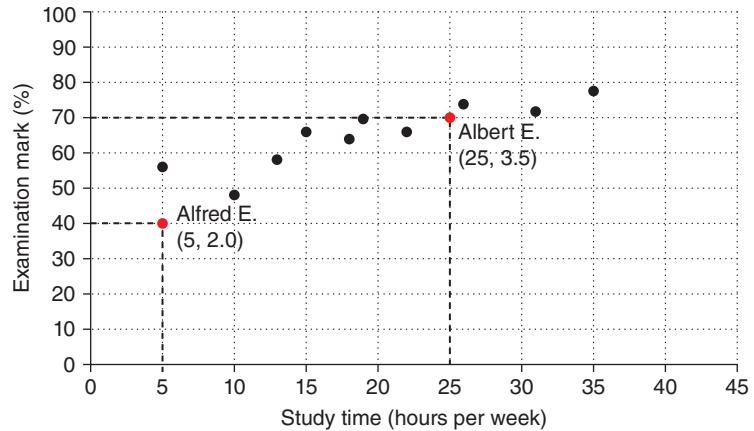
Suppose you want to examine the relationship between study time and examination marks for students attending economics lectures. You could record a pair of numbers: hours per week spent studying and marks obtained in the final course examination. These numbers could then be placed in parentheses as an *ordered pair* and appear as a single point on the graph. Albert, for instance, is represented by the ordered pair (25 hours/week, 70 per cent examination mark), while his classmate Alfred is represented by the ordered pair (5 hours/week, 40 per cent examination mark).

We can graph these ordered pairs on a two-dimensional grid. The first number in each ordered pair, called the *x-coordinate*, tells us the horizontal location of the point. The second number, called the *y-coordinate*, tells us the vertical location of the point. The point with both an *x-coordinate* and a *y-coordinate* of zero is known as the *origin*. The two coordinates in the ordered pair tell us where the point is located in relation to the origin: *x* units to the right of the origin and *y* units above it.

Figure 2A.3 graphs examination marks against study time for Albert, Alfred and the rest of the students who attended the course. This type of graph is called a *scatterplot* because it plots scattered points. Looking at this graph, we immediately notice that points further to the right (indicating more study time) also tend to be higher (indicating a better examination result). Because study time and examination mark typically move in the same direction, we say that these two variables have a *positive correlation*. By contrast, if we were to graph time spent partying per week and examination marks, we would probably find that higher party time is associated with lower marks; because these variables typically move in opposite directions, we would call this a *negative correlation*. In either case, the coordinate system makes the correlation between the two variables easy to see.

FIGURE 2A.3**Using the Coordinate System**

Final examination mark is measured on the vertical axis and study time on the horizontal axis. Albert, Alfred and the other students on their course are represented by various points. We can see from the graph that students who spend more hours studying tend to get higher marks.

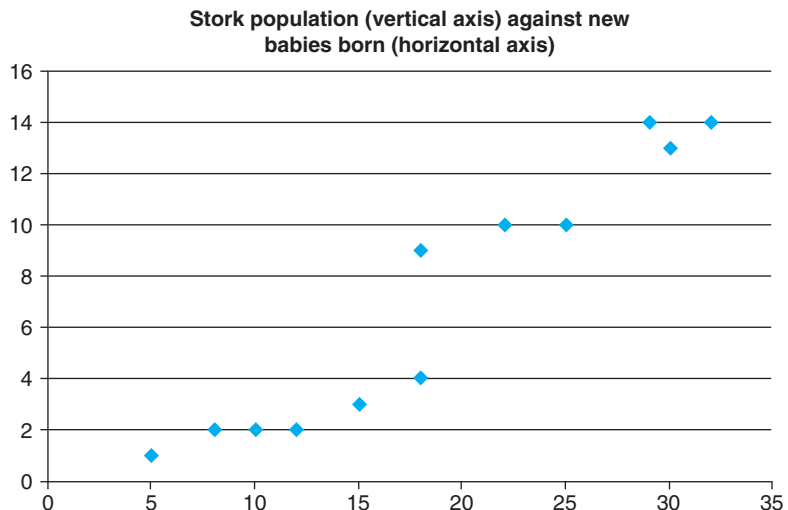


Limitations of Scatter Plot Graphs Looking at the scatter plot in Figure 2A.3, it seems intuitively sensible that more time devoted to study is associated with higher marks. But just because two variables appear to have a relationship does not mean we can simply accept the data at face value. As economists we have to be critical – we have to question what we see. The following example highlights this issue.

Assume that data are released in a particular area which we shall call Region X, which shows a rise in the number of new born babies over a period of time. At the same time, it is observed that the stork population in Region X has also risen over the same period of time. There is a ‘theory’ that storks bring babies. This ‘theory’ comes from the belief that storks are a fertility symbol. Storks live in and around marsh and wetland areas. There was also an ancient belief that the souls of unborn babies lived in water and that storks carried these souls to mothers – hence the belief that storks bring babies.

FIGURE 2A.4**The Stork Population and New Born Babies in Region X**

The scatter plot of the information in the Table shows the stork population over a period of time in Region X and the number of new born babies over the same time period.



Looking at the scatter plot in Figure 2A.4 it seems clear that there is a correlation between the two variables and one might be drawn to the conclusion that storks do indeed bring babies. However, common sense and biology tell us that what we are seeing is simply a coincidence – two variables which seem to have some correlation but where the correlation is due to chance. To determine whether there is a relationship between variables and the extent to which any relationship is due to chance can be identified by the

use of different statistical tests. If the appropriate tests were carried out on this data it is likely to tell us that there is a significant correlation and the extent to which the correlation is due to chance. This example serves as another reminder that economists must always question and never simply accept things at face value.

Curves in the Coordinate System

Students who study more do tend to get higher marks, but other factors also influence a student's marks. Previous preparation is an important factor, for instance, as are talent, attention from teachers and even eating a good breakfast. A scatter plot like Figure 2A.3 does not attempt to isolate the effect that study has on grades from the effects of other variables. Often, however, economists prefer looking at how one variable affects another while holding everything else constant. The principle used here is referred to as *ceteris paribus*, a Latin phrase which roughly translated means 'other things being equal'. Economists analyse problems which they fully understand have multiple factors which could be affecting outcomes being investigated. In order to try and isolate the factors which are the most significant ones, the principle of *ceteris paribus* is used. It enables us to look at how changes in one variable affect outcomes assuming other variables that we know might have an effect are held constant. We can then look at the other variables in turn and build a more complete picture of the effect of changes in variables.

To see how this is done, let's consider one of the most important graphs in economics – the *demand curve*. The demand curve traces out the effect of a good's price on the quantity of the good consumers want to buy. Before showing a demand curve, however, consider Table 2A.1, which shows how the number of novels that Maria buys depends on her income and on the price of novels. When novels are cheap, Maria buys them in large quantities. As they become more expensive, she borrows books from the library instead of buying them or chooses to go to the cinema instead of reading. Similarly, at any given price, Maria buys more novels when she has a higher income. That is, when her income increases, she spends part of the additional income on novels and part on other goods.

TABLE 2A.1 Novels Purchased by Maria

This table shows the number of novels Maria buys at various incomes and prices. For any given level of income, the data on price and quantity demanded can be graphed to produce Maria's demand curve for novels, as shown in Figures 2A.5 and 2A.6.

| Price | Income | | |
|-------|------------------------------------|------------------------------------|------------------------------------|
| | €20,000 | €30,000 | €40,000 |
| €10 | 2 novels | 5 novels | 8 novels |
| €9 | 6 | 9 | 12 |
| €8 | 10 | 13 | 16 |
| €7 | 14 | 17 | 20 |
| €6 | 18 | 21 | 24 |
| €5 | 22 | 25 | 28 |
| | <i>Demand curve, D₃</i> | <i>Demand curve, D₁</i> | <i>Demand curve, D₂</i> |

We now have three variables – the price of novels, income and the number of novels purchased – which is more than we can represent in two dimensions. To put the information from Table 2A.1 in graphical form, we need to hold one of the three variables constant and trace out the relationship between the other two. Because the demand curve represents the relationship between price and quantity demanded, we hold Maria's income constant and show how the number of novels she buys varies with the price of novels.

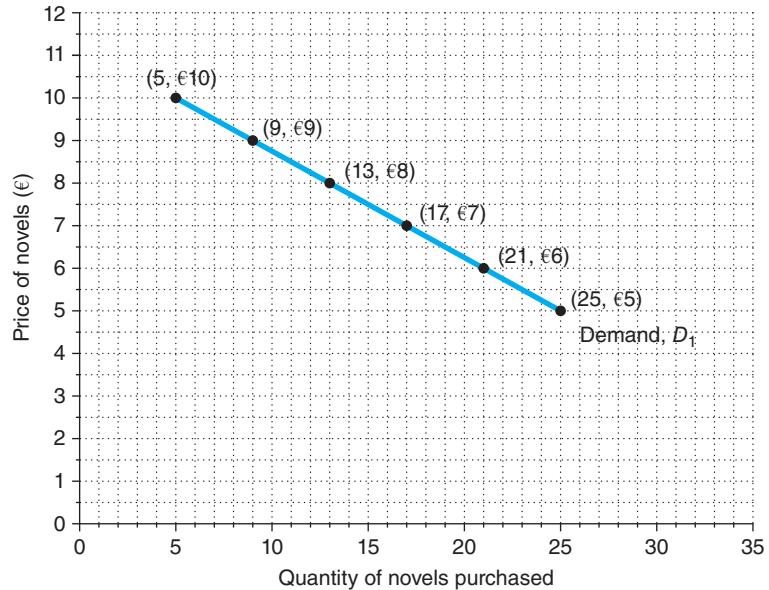
Suppose that Maria's income is €30,000 per year. If we place the number of novels Maria purchases on the *x*-axis and the price of novels on the *y*-axis, we can graphically represent the middle column of Table 2A.1. When the points that represent these entries from the table – (5 novels, €10), (9 novels, €9) and so on – are connected, they form a line. This line, pictured in Figure 2A.5, is known as Maria's demand curve for novels; it tells us how many novels Maria purchases at any given price. The demand curve is

downward sloping, indicating that a higher price reduces the quantity of novels demanded. Because the quantity of novels demanded and the price move in opposite directions, we say that the two variables are *negatively or inversely related*. (Conversely, when two variables move in the same direction, the curve relating them is upward sloping, and we say the variables are *positively related*.)

FIGURE 2A.5

Demand Curve

The line D_1 shows how Maria's purchases of novels depend on the price of novels when her income is held constant. Because the price and the quantity demanded are negatively related, the demand curve slopes downward.



Now suppose that Maria's income rises to €40,000 per year. At any given price, Maria will purchase more novels than she did at her previous level of income. Just as earlier we drew Maria's demand curve for novels using the entries from the middle column of Table 2A.1, we now draw a new demand curve using the entries from the right-hand column of the table. This new demand curve (curve D_2) is pictured alongside the old one (curve D_1) in Figure 2A.6; the new curve is a similar line drawn farther to the right. We therefore say that Maria's demand curve for novels *shifts to the right* when her income increases. Likewise, if Maria's income were to fall to €20,000 per year, she would buy fewer novels at any given price and her demand curve would shift to the left (to curve D_3).

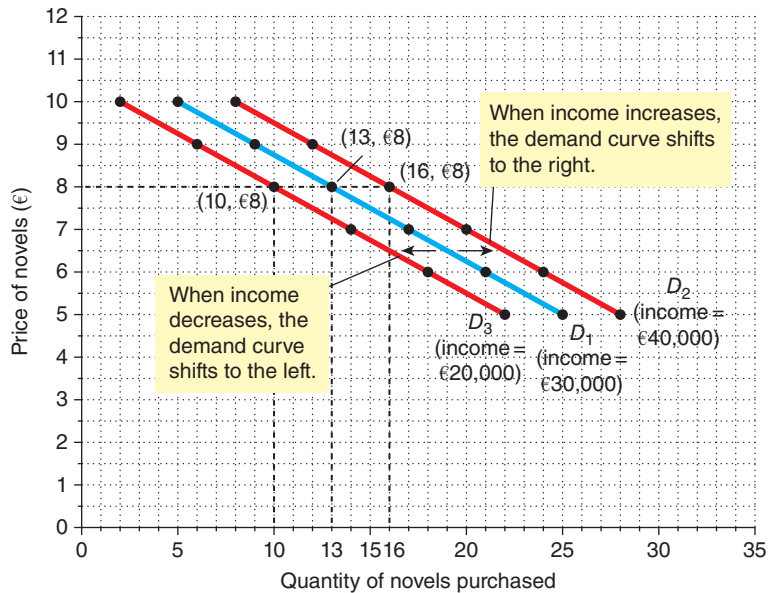
In economics, it is important to distinguish between *movements along a curve* and *shifts of a curve*. As we can see from Figure 2A.5, if Maria earns €30,000 per year and novels are priced at €8 apiece, she will purchase 13 novels per year. If the price of novels falls to €7, Maria will increase her purchases of novels to 17 per year. The demand curve, however, stays fixed in the same place. Maria still buys the same number of novels *at each price*, but as the price falls she moves along her demand curve from left to right. By contrast, if the price of novels remains fixed at €8 but her income rises to €40,000, Maria increases her purchases of novels from 13 to 16 per year. Because Maria buys more novels *at each price*, her demand curve shifts out, as shown in Figure 2A.6.

There is a simple way to tell when it is necessary to shift a curve. When a variable that is not named on either axis changes, the curve shifts. Income is on neither the x -axis nor the y -axis of the graph, so when Maria's income changes, her demand curve must shift. Any change that affects Maria's purchasing habits, besides a change in the price of novels, will result in a shift in her demand curve. If, for instance, the public library closes and Maria must buy all the books she wants to read, she will demand more novels at each price, and her demand curve will shift to the right. Or, if the price of going to the cinema falls and Maria spends more time at the movies and less time reading, she will demand fewer novels at each price, and her demand curve will shift to the left. By contrast, when a variable on an axis of the graph changes, the curve does not shift. We read the change as a movement along the curve.

FIGURE 2A.6

Shifting Demand Curves

The location of Maria's demand curve for novels depends on how much income she earns. The more she earns, the more novels she will purchase at any given price, and the further to the right her demand curve will lie. Curve D_1 represents Maria's original demand curve when her income is €30,000 per year. If her income rises to €40,000 per year, her demand curve shifts to D_2 . If her income falls to €20,000 per year, her demand curve shifts to D_3 .

**Slope**

One question we might want to ask about Maria is how much her purchasing habits respond to price. Look at the demand curve D_2 pictured in Figure 2A.7. This curve is very steep, Maria purchases nearly the same number of novels regardless of whether they are cheap or expensive. The demand curve D_1 is much flatter, Maria purchases fewer novels when the price rises. To answer questions about how much one variable responds to changes in another variable, we can use the concept of *slope*.

The slope of a line is the ratio of the vertical distance covered to the horizontal distance covered as we move along the line. This definition is usually written out in mathematical symbols as follows:

$$\text{slope} = \frac{\Delta y}{\Delta x}$$

where the Greek letter Δ (delta) stands for the change in a variable. In other words, the slope of a line is equal to the 'rise' (change in y) divided by the 'run' (change in x). The slope will be a small positive number for a fairly flat upward sloping line, a large positive number for a steep upward sloping line and a negative number for a downward sloping line. A horizontal line has a slope of zero because in this case the y -variable never changes; a vertical line is said to have an infinite slope because the y -variable can take any value without the x -variable changing at all.

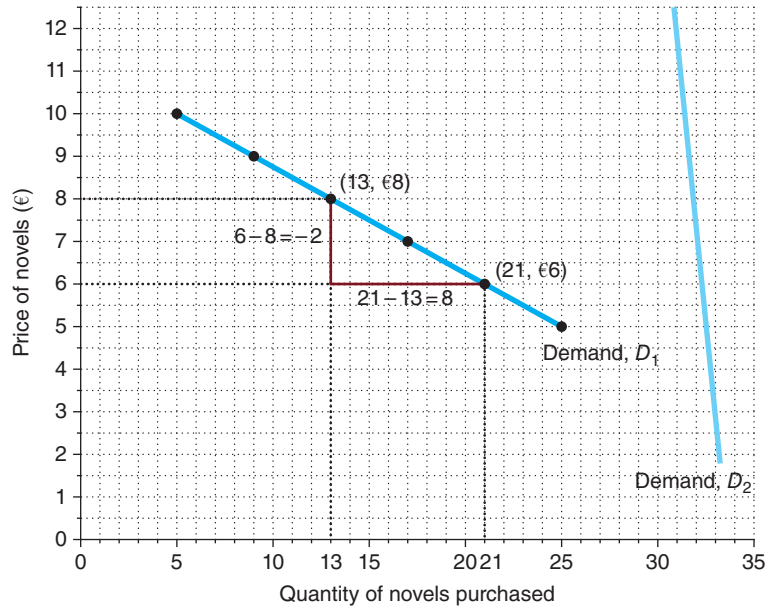
What is the slope of Maria's demand curve for novels? First of all, because the curve slopes down, we know the slope will be negative. To calculate a numerical value for the slope, we must choose two points on the line. With Maria's income at €30,000, she will purchase 21 novels at a price of €6 or 13 novels at a price of €8. When we apply the slope formula, we are concerned with the change between these two points; in other words, we are concerned with the difference between them, which lets us know that we will have to subtract one set of values from the other, as follows:

$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{\text{first } y\text{-coordinate} - \text{second } y\text{-coordinate}}{\text{first } x\text{-coordinate} - \text{second } x\text{-coordinate}} = \frac{6 - 8}{21 - 13} = \frac{-2}{8} = \frac{-1}{4}$$

Figure 2A.7 shows graphically how this calculation works. Try computing the slope of Maria's demand curve using two different points. You should get exactly the same result, $-1/4$. One of the properties of a straight line is that it has the same slope everywhere. This is not true of other types of curves, which are steeper in some places than in others.

FIGURE 2A.7**Calculating the Slope of a Line**

To calculate the slope of the demand curve, we can look at the changes in the y - and x -coordinates as we move from the point (21 novels, €6) to the point (13 novels, €8). The slope of the line is the ratio of the change in the y -coordinate (-2) to the change in the x -coordinate ($+8$), which equals $-1/4$.



The slope of Maria's demand curve tells us something about how responsive her purchases are to changes in the price. A small slope (a number close to zero) means that Maria's demand curve is relatively flat; in this case, she adjusts the number of novels she buys substantially in response to a price change. A larger slope (a number further from zero) means that Maria's demand curve is relatively steep; in this case she adjusts the number of novels she buys only slightly in response to a price change.

Cause and Effect

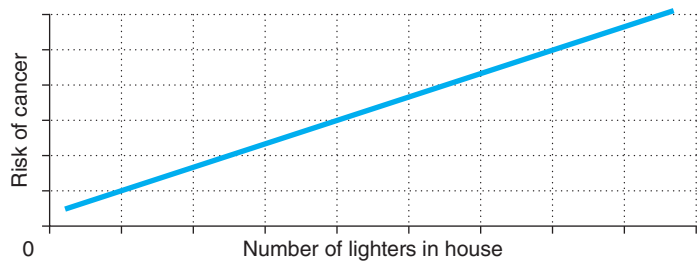
Economists often use graphs to advance an argument about how the economy works. In other words, they use graphs to argue about how one set of events *causes* another set of events. With a graph like the demand curve, there is no doubt about cause and effect. Because we are varying price and holding all other variables constant, we know that changes in the price of novels cause changes in the quantity Maria demands. Remember, however, that our demand curve came from a hypothetical example. When graphing data from the real world, it is often more difficult to establish how one variable affects another.

The first problem is that it is difficult to hold everything else constant when measuring how one variable affects another. If we are not able to hold variables constant, we might decide that one variable on our graph is causing changes in the other variable when actually those changes are caused by a third *omitted variable* not pictured on the graph. Even if we have correctly identified the two variables to look at, we might run into a second problem – *reverse causality*. In other words, we might decide that A causes B when in fact B causes A. The omitted variable and reverse causality traps require us to proceed with caution when using graphs to draw conclusions about causes and effects.

Omitted Variables To see how omitting a variable can lead to a deceptive graph, let's consider an example. Imagine that the government, spurred by public concern about the large number of deaths from cancer, commissions an exhaustive study from Big Brother Statistical Services. Big Brother examines many of the items found in people's homes to see which of them are associated with the risk of cancer. Big Brother reports a strong relationship between two variables: the number of cigarette lighters that a household owns and the probability that someone in the household will develop cancer. Figure 2A.8 shows this relationship.

FIGURE 2A.8**Graph With an Omitted Variable**

The upward sloping curve shows that members of households with more cigarette lighters are more likely to develop cancer. Yet we should not conclude that ownership of lighters causes cancer, because the graph does not take into account the number of cigarettes smoked.



What should we make of this result? Big Brother advises a quick policy response. It recommends that the government discourages the ownership of cigarette lighters by taxing their sale. It also recommends that the government requires warning labels: 'Big Brother has determined that this lighter is dangerous to your health.'

In judging the validity of Big Brother's analysis, one question is paramount: has Big Brother held constant every relevant variable except the one under consideration? If the answer is no, the results are suspect. An easy explanation for Figure 2A.8 is that people who own more cigarette lighters are more likely to smoke cigarettes and that cigarettes, not lighters, cause cancer. If Figure 2A.8 does not hold constant the amount of smoking, it does not tell us the true effect of owning a cigarette lighter.

This story illustrates an important principle: when you see a graph being used to support an argument about cause and effect, it is important to ask whether the movements of an omitted variable could explain the results you see.

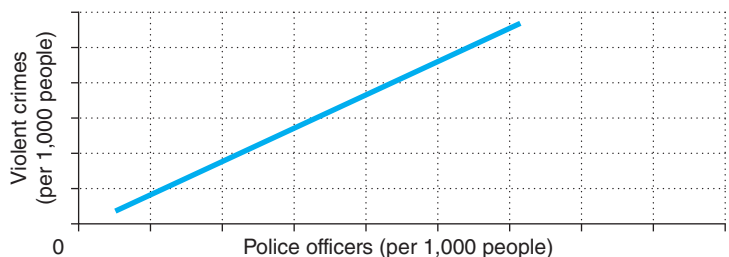
Reverse Causality Economists can also make mistakes about causality by misreading its direction. To see how this is possible, suppose the Association of European Anarchists commissions a study of crime in Eurovia and arrives at Figure 2A.9, which plots the number of violent crimes per 1,000 people in major Eurovian cities against the number of police officers per 1,000 people. The anarchists note the curve's upward slope and argue that because police increase rather than decrease the amount of urban violence, law enforcement should be abolished.

If we could run a controlled experiment, we would avoid the danger of reverse causality. To run an experiment, we would set the number of police officers in different cities randomly and then examine the correlation between police and crime. Figure 2A.9, however, is not based on such an experiment. We simply observe that more dangerous cities have more police officers. The explanation for this may be that more dangerous cities hire more police. In other words, rather than police causing crime, crime may cause police. Nothing in the graph itself allows us to establish the direction of causality.

It might seem that an easy way to determine the direction of causality is to examine which variable moves first. If we see crime increase and then the police force expand, we reach one conclusion. If we see the police force expand and then crime increase, we reach the other. Yet there is also a flaw with this approach: often people change their behaviour not in response to a change in their present conditions but

FIGURE 2A.9**Graph Suggesting Reverse Causality**

The upward sloping curve shows that Eurovian cities with a higher concentration of police are more dangerous. Yet the graph does not tell us whether police cause crime or crime-plagued cities hire more police.



in response to a change in their *expectations* of future conditions. A city that expects a major crime wave in the future, for instance, might well hire more police now. This problem is even easier to see in the case of babies and baby cots. Couples often buy a baby cot in anticipation of the birth of a child. The cot comes before the baby, but we wouldn't want to conclude that the sale of cots causes the population to grow!

There is no complete set of rules that says when it is appropriate to draw causal conclusions from graphs. Yet just keeping in mind that cigarette lighters don't cause cancer (omitted variable) and baby cots do not cause larger families (reverse causality) will keep you from falling for many faulty economic arguments.

Constrained Optimization

Throughout this book we will look at situations where consumers, firms and governments are assumed to exhibit maximizing or minimizing behaviour. Consumers might be assumed to want to maximize the satisfaction (utility) they get from consumption, firms might be assumed to want to maximize profits but minimize costs, and governments might want to maximize tax revenue, for example. In most cases, there will be factors which will limit the extent to which this maximizing or minimizing behaviour can be carried out such as time, income and resources. In the case of consumers, the ability to maximize utility will be subject to a constraint – their income. Firms might want to minimize costs but subject to the constraint of their income (revenue) or the factor inputs that they have at their disposal.

Economists will often carry out analysis under conditions of constrained optimization. Typically this might take the form: maximize x subject to the existence of y constraint. Usually, the constraint is assumed to be fixed and so a calculation can be made which shows the optimizing (maximum or minimum) behaviour given the constraint which exists. We can then change the constraint to see how behaviour would differ and what the outcome would be. You will probably learn some of the mathematical techniques for solving constrained optimization problems in the quantitative methods modules that you will take alongside your Principles of Economics module.

Real versus Nominal Values

Economists deal with numbers but it is important to have some understanding of the numbers we might happen to be working with. Distinguishing between real and nominal values is extremely important. A simple way to understand the difference between nominal and real values is to remember that nominal values are expressed in money terms whereas real values are expressed as quantities. Typically, we will look at values which are affected by price movements over time. For example, if a firm was to tell you that it had sold €1 million worth of extra products over the last year (2014) what does that tell you? The answer is: not very much! If the firm sold €10 million worth of goods the year before (2013) you might be tempted to think that now selling €11 million worth is a 'good' thing and indeed, it might be.

Now assume that you know a little bit more about the situation of the firm in our example. Assume that the firm sold 10 million units in 2013, each priced at €1. Making a judgement on the performance of the firm in 2014 requires we know a little bit more about the extra €1 million worth of sales – the nominal value. If the price of the goods sold in 2014 stayed at €1 then we know that the firm sold an extra 1 million units that year – a 10 per cent increase in sales volumes, which we might conclude is a creditable performance. However, would you come to the same conclusion if you discovered that the price they sold their products at in 2014 was €10, meaning they only sold an extra 100,000 units in that year? Now the increase in sales is only 1 per cent – not such a good performance. The performance of the firm is largely due to the change in price not the amount of goods sold. In this case, nominal sales rose by €1 million but real sales rose by 100,000 units.

Real values, therefore, take into account price changes over time or the adjusting factors such as the seasons, whereas nominal values don't take account of these adjusting factors. For this reason you might see nominal values referred to as *current prices* whereas real values might be referred to as *constant prices*.

PART 2

SUPPLY AND DEMAND: HOW MARKETS WORK

3 THE MARKET FORCES OF SUPPLY AND DEMAND

In April 2012, the price of winter wheat on global markets was around €209 per metric tonne. By September 2012 the price had risen to €277 per metric tonne, a rise of around 33 per cent. One of the reasons for the rise in price was that wheat-producing states in the United States suffered a drought which damaged the wheat crop. Other crops such as corn and soya beans had also been rising in price. One of the consequences of this is that livestock farmers faced increased costs of production – some arable crops are used in livestock feedstuffs. Some farmers had been unable to afford to feed their animals and so had to send them for slaughter which led to beef prices falling.

In November 2006, the average house price in the UK was £200,000 (€250,000). In January 2009, this had fallen to £164,000 (€205,000). When major sporting events are held in a country, such as the World Cup or the Olympics, hotel prices in the areas around the venues tend to rise. In the Holy month of Ramadan, many Muslims find that the price of food rises quite sharply. If you book an airline flight several months in advance of a trip the seat prices tend to be lower than if you try booking two weeks before your trip. What do these events have in common? They all show the workings of supply and demand.

Supply and *demand* are the two words that economists use most often – and for good reason. Supply and demand are the forces that make market economies work. They determine the quantity of each good produced and the price at which it is sold. If you want to know how any event or policy will affect the economy, you must think first about how it will affect supply and demand.

This chapter introduces the theory of supply and demand. It considers how buyers and sellers behave and how they interact with one another. It shows how supply and demand determine prices in a market economy and how prices, in turn, allocate the economy's scarce resources.

MARKETS AND COMPETITION

The terms *supply* and *demand* refer to the behaviour of people as they interact with one another in markets. A **market** is a group of buyers and sellers of a particular good or service. The buyers as a group determine the demand for the product, and the sellers as a group determine the supply of the product. Before discussing how buyers and sellers behave, let's first consider more fully what we mean by a 'market' and the various types of markets we observe in the economy.

market a group of buyers and sellers of a particular good or service

Competitive Markets

Markets take many forms. Sometimes markets are highly organized, such as the markets for many agricultural commodities and for metals. In these markets, buyers and sellers meet at a specific time and place, where an auctioneer helps set prices and arrange sales.

More often, markets are less organized. For example, consider the market for coffee in a particular town. Buyers of coffee do not meet together at any one time. The sellers of coffee are in different locations and offer different products (lattes, Americano, espresso, instant, filter, etc.). There is no auctioneer calling out the price of coffee. Each seller posts a price for a cup of coffee in their shop, and each buyer decides how much coffee to buy at each shop.

Even though it is not organized, the group of coffee buyers and coffee sellers forms a market. Each buyer knows that there are several sellers from which to choose, and each seller is aware that his product is similar to that offered by other sellers. The price of coffee and the quantity of coffee sold are not determined by any single buyer or seller. Rather, price and quantity are determined by all buyers and sellers as they interact in the marketplace.

The market for coffee, like most markets in the economy, is highly competitive. Competition exists when two or more firms are rivals for customers. Each firm strives to gain the attention and custom of buyers in the market. The market for coffee in a town will consist of established chains which are familiar across Europe – Costa, Starbucks, Café Nero, Koffie Café, Lavazza and Coffeeheaven, for example – but there will also be many small independent cafés, bars and restaurants where buyers can get a cup of coffee. Any of these coffee sellers faces competition from other coffee sellers.

In economics, however, there are particular characteristics ascribed to the term ‘competitive market’. A **competitive market** is a market in which there are many buyers and many sellers so that each has a negligible impact on the market price. Each seller has limited control over the price because other sellers are offering similar products and each seller only supplies a very small amount in relation to the total supply of the market. In a perfectly competitive market the products are identical (homogenous) so a seller has little reason to charge less than the going price, and if he or she charges more, buyers will make their purchases elsewhere. Similarly, no single buyer can influence the price because each buyer purchases only a small amount relative to the size of the market.

In this chapter, we will look at the model for supply and demand assuming these characteristics; we examine how buyers and sellers interact in competitive markets. We see how the forces of supply and demand determine both the quantity of the good sold and its price. In later chapters we will look at what happens when some of these assumptions are relaxed. By doing this we can see how markets might work differently when there are a few sellers who dominate the market or when they are able to differentiate their product in some way and so the assumption of homogeneity does not hold.

competitive market a market in which there are many buyers and sellers so that each has a negligible impact on the market price

Competition: Perfect and Otherwise

The assumptions of perfect competition outlined above lead us to some important conclusions. Because there are many buyers and sellers in a perfectly competitive market, neither has any power to influence price – they must accept the price the market determines, they are said to be *price takers*.

There are some markets in which the assumption of perfect competition applies to a very large degree. In the EU agriculture market, for example, there are about 14 million farmers who sell cereals, fruit, milk, beef, lamb and so on, and millions of consumers who buy these products. Because no single buyer or seller can influence the price of agricultural products, each takes the price as given.

Not all goods and services, however, are sold in perfectly competitive markets. Some markets have only one seller, and this seller sets the price. Such a seller is called a *monopoly*. Your local water company, for instance, may be a monopoly. Residents in your area probably have only one water company from which to buy this service.

Some markets fall between the extremes of perfect competition and monopoly. One such market, called an *oligopoly*, has a few sellers that do not always compete aggressively. There are many examples of oligopolistic markets across Europe in products like steel, tyre manufacture, retailing (supermarkets), alcoholic beverages, telecommunications, pharmaceuticals, banking and athletic sportswear. Another type of market is *monopolistically competitive*; it contains many sellers but each offers a slightly different product. Because the products are not exactly the same, each seller has some ability to set the price for its own product. An example is the market for magazines. Magazines compete with one another for readers and anyone can enter the market by starting a new one, but each magazine offers different articles and can set its own price.

Despite the diversity of market types we find in the world, we begin by studying perfect competition. Perfectly competitive markets are the easiest to analyse. Moreover, because some degree of competition is present in most markets, many of the lessons that we learn by studying supply and demand under perfect competition apply in more complicated markets as well.

SELF TEST What constitutes a market? List the main characteristics of a competitive market.

DEMAND

We begin our study of markets by examining the behaviour of buyers. To focus our thinking, let's keep in mind a particular good – milk. The market for milk fulfils many of the characteristics of a perfectly competitive market; milk is fairly homogenous, there are about half a million dairy farms across the EU and there are millions of buyers of milk across the EU.

The Demand Curve: The Relationship Between Price and Quantity Demanded

The **quantity demanded** of any good is the amount of the good that buyers are willing and able to purchase at different prices. Many things determine the quantity demanded of any good, but when analysing how markets work, one determinant plays a central role – the price of the good. If the price of milk rose from €0.25 per litre to €0.35 per litre, less milk would be bought. If the price of milk fell to €0.20 per litre more milk would be bought. Because the quantity demanded falls as the price rises and rises as the price falls, we say that the quantity demanded is *negatively or inversely related* to the price. This relationship between price and quantity demanded is true for most goods in the economy and, in fact, is so pervasive that economists call it the **law of demand**: other things equal, when the price of a good rises, the quantity demanded of the good falls, and when the price falls, the quantity demanded rises.

quantity demanded the amount of a good that buyers are willing and able to purchase at different prices
law of demand the claim that, other things equal (*ceteris paribus*) the quantity demanded of a good falls when the price of the good rises

We can represent the relationship between the price and quantity demanded in a table such as the one shown in Figure 3.1. The table shows how many litres of milk Rachel is willing and able to buy each month at different prices for milk. If milk is free, Rachel would be willing to buy 20 litres. At €0.10 per litre, Rachel would be willing to buy 18 litres. As the price rises further, she is willing to buy fewer and fewer litres. When the price reaches €1, Rachel would not be prepared to buy any milk at all. This table is a **demand schedule**, a table that shows the relationship between the price of a good and the quantity demanded, holding constant everything else that influences how much consumers of the good want to buy.

demand schedule a table that shows the relationship between the price of a good and the quantity demanded

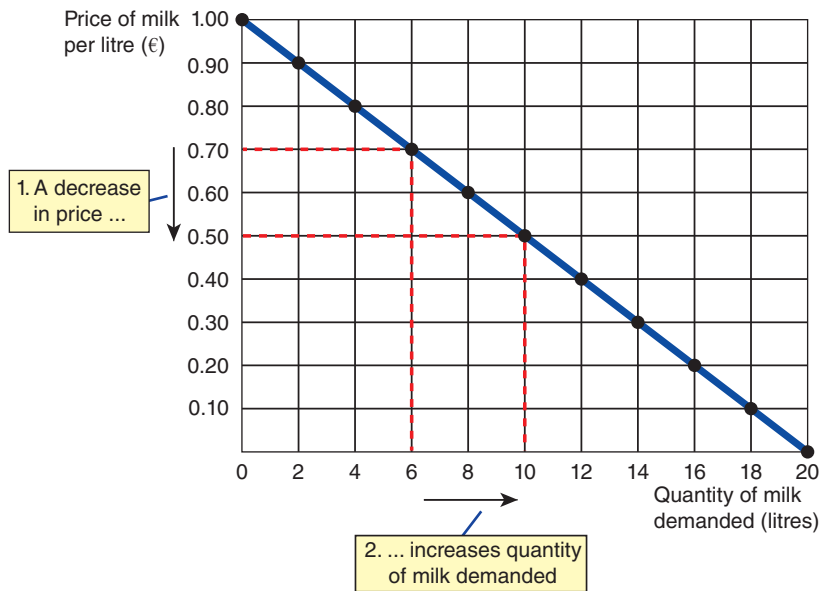
The graph in Figure 3.1 uses the numbers from the table to illustrate the law of demand. By convention, price is on the vertical axis, and the quantity demanded is on the horizontal axis. The downwards sloping line relating price and quantity demanded is called the **demand curve**.

demand curve a graph of the relationship between the price of a good and the quantity demanded

FIGURE 3.1

Rachel's Demand Schedule and Demand Curve

The demand schedule shows the quantity demanded at each price. The demand curve, which graphs the demand schedule, shows how the quantity demanded of the good changes as its price varies. Because a lower price increases the quantity demanded, the demand curve slopes downwards.



| Price of milk per litre (€) | Quantity of milk demanded (litres per month) |
|-----------------------------|--|
| 0.00 | 20 |
| 0.10 | 18 |
| 0.20 | 16 |
| 0.30 | 14 |
| 0.40 | 12 |
| 0.50 | 10 |
| 0.60 | 8 |
| 0.70 | 6 |
| 0.80 | 4 |
| 0.90 | 2 |
| 1.00 | 0 |

Market Demand versus Individual Demand

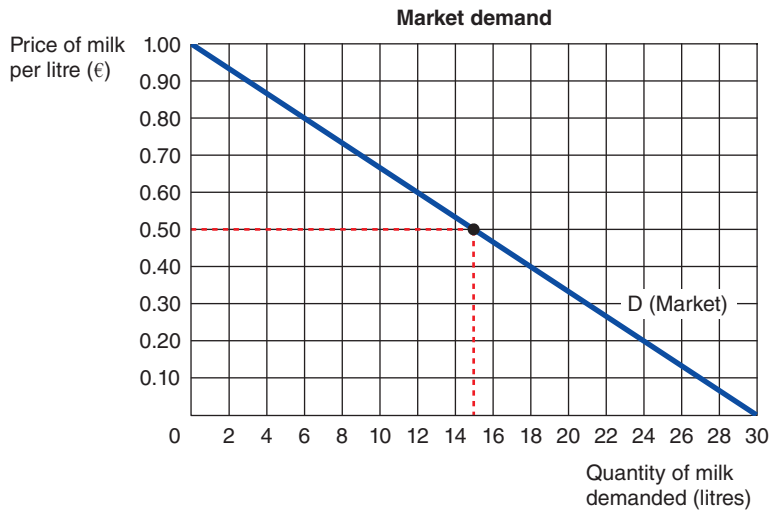
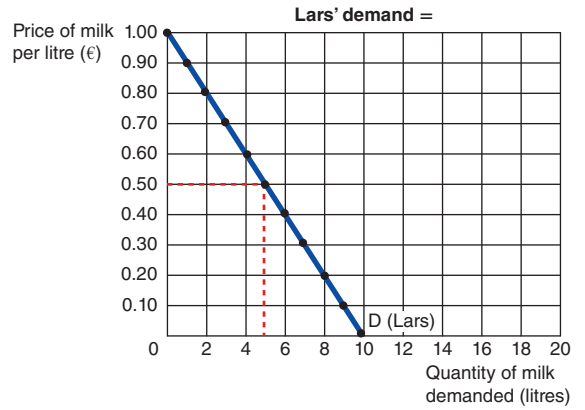
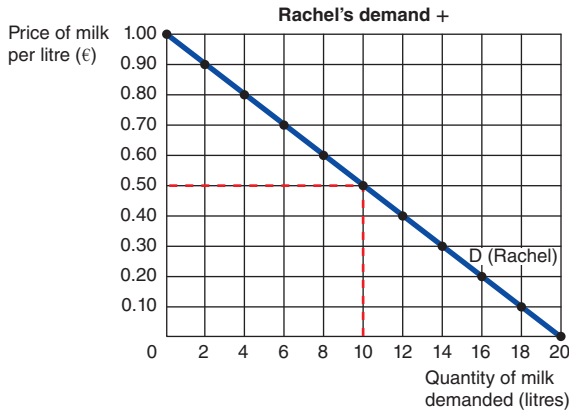
The demand curve in Figure 3.1 shows an individual's demand for a product. To analyse how markets work, we need to determine the *market demand*, which is the sum of all the individual demands for a particular good or service.

The table in Figure 3.2 shows the demand schedules for milk of two individuals – Rachel and Lars. At any price, Rachel's demand schedule tells us how much milk she would like to buy and Lars' demand

FIGURE 3.2

Market Demand as the Sum of Individual Demands

The quantity demanded in a market is the sum of the quantities demanded by all the buyers at each price. Thus, the market demand curve is found by adding horizontally the individual demand curves. At a price of €0.50, Rachel would like to buy 10 litres of milk but Lars would only be prepared to buy 5 litres at that price. The quantity demanded **in the market** at this price, therefore, is 15 litres.



| Price of milk per litre (€) | Rachel + | Lars = | Market |
|-----------------------------|----------|--------|--------|
| 0.00 | 20 | 10 | 30 |
| 0.10 | 18 | 9 | 27 |
| 0.20 | 16 | 8 | 24 |
| 0.30 | 14 | 7 | 21 |
| 0.40 | 12 | 6 | 18 |
| 0.50 | 10 | 5 | 15 |
| 0.60 | 8 | 4 | 12 |
| 0.70 | 6 | 3 | 9 |
| 0.80 | 4 | 2 | 6 |
| 0.90 | 2 | 1 | 3 |
| 1.00 | 0 | 0 | 0 |

schedule tells us how much milk he is willing to buy. The market demand at each price is the sum of the two individual demands.

Figure 3.2 shows the demand curves that correspond to these demand schedules. To find the total quantity demanded at any price, we add the individual quantities found on the horizontal axis of the individual demand curves. The market demand curve shows how the total quantity demanded of a good varies as the price of the good varies, while all the other factors that affect how much consumers want to buy, such as incomes and taste, among other things, are held constant.

Shifts versus Movements along the Demand Curve

One important distinction that must be made is between a shift of the demand curve and a movement along the demand curve. A shift in the demand curve is caused by a factor affecting demand **other than a change in price**. The factors affecting demand are outlined below presented as a function written:

$$D = f(P_n, P_1, P_2, \dots, P_{n-1}, Y, T, P_{LS}, A, E)$$

Where:

- P_n = Price
- P_1, P_2, \dots, P_{n-1} = Prices of other goods – substitutes and complements
- Y = Incomes – the level and distribution of income
- T = Tastes and fashions
- P_{LS} = The level and structure of the population
- A = Advertising
- E = Expectations of consumers

If any of the variables affecting demand *apart from* price changes then the demand curve will shift, either to the right (an increase in demand) or to the left (a reduction in demand). For example, if the price of milk is €0.30 per litre a family might buy 5 litres of milk a week. If their income rises, they can now afford to buy more milk and so might now buy 7 litres a week. The price of milk has not changed – it is still €0.30 per litre but the amount of milk the family buys has increased.

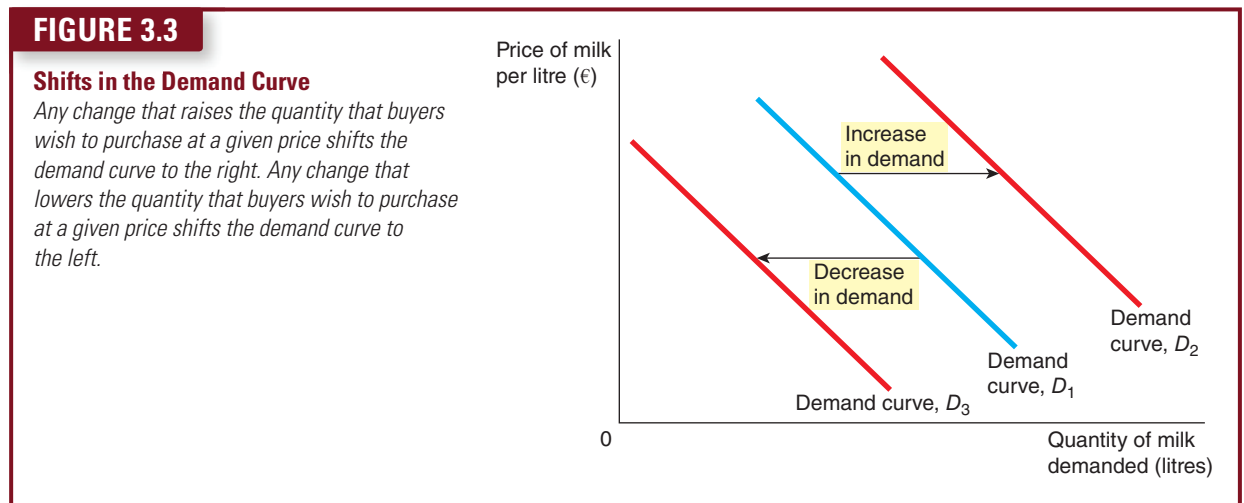
If any of the factors affecting demand other than price change then the amount consumers wish to purchase changes whatever the price. The shift in the demand curve is referred to as an *increase or decrease in demand*. A movement along the demand curve occurs when there is a change in price. This may occur because of a change in supply conditions. A change in price leads to a movement along the demand curve and is referred to as a *change in quantity demanded*.

Movement Along the Demand Curve Let us assume that the price of milk falls. We know that the fall in price will lead to an increase in quantity demanded. There are two reasons for this increase:

1. **The income effect.** If we assume that incomes remain constant then a fall in the price of milk means that consumers can now afford to buy more with their income. In other words, their *real income*, what a given amount of money can buy at any point in time, has increased and part of the increase in quantity demanded can be put down to this effect.
2. **The substitution effect.** Now that milk is lower in price compared to other products such as fruit juice, some consumers will choose to substitute the more expensive drinks with the now cheaper milk. This switch accounts for the remaining part of the increase in quantity demanded.

A Shift in the Demand Curve The demand curve for milk shows how much milk people are willing to buy at any given price, holding constant the many other factors beyond price that influence consumers' buying decisions. As a result, this demand curve need not be stable over time. If something happens to alter the quantity demanded at any given price, the demand curve shifts. For example, suppose a top European medical school published research findings that suggested people who regularly drink milk live longer, healthier lives. The discovery would raise the demand for milk. At any given price, buyers would now want to purchase a larger quantity of milk and the demand curve for milk would shift.

Figure 3.3 illustrates shifts in demand. Any change that increases the quantity demanded at every price, such as our imaginary research report, shifts the demand curve to the right and is called *an increase in demand*. Any change that reduces the quantity demanded at every price shifts the demand curve to the left and is called *a decrease in demand*.



The following is a short summary of the factors affecting demand, changes in which cause a shift of the demand curve.

Prices of Other (Related) Goods Suppose that the price of milk falls. The law of demand says that you will buy more milk. At the same time, you will probably buy less fruit juice. Because milk and fruit juice are both refreshing drinks, they satisfy similar desires. When a fall in the price of one good reduces the demand for another good, the two goods are called **substitutes**. Substitutes are often pairs of goods that are used in place of each other, such as butter and margarine, pullovers and sweatshirts, and cinema tickets and DVD rentals. The more closely related substitute products are, the more effect we might see on demand if the price of one of the substitutes changes.

substitutes two goods for which an increase in the price of one leads to an increase in the demand for the other

Now suppose that the price of breakfast cereals falls. According to the law of demand, more packets of breakfast cereals will be bought. Yet, in this case, the demand for milk will increase as well, because breakfast cereals and milk are used together. When a fall in the price of one good raises the demand for another good, the two goods are called **complements**. Complements are often pairs of goods that are used together, such as petrol and cars, computers and software, bread and cheese, strawberries and cream, and bacon and eggs.

complements two goods for which an increase in the price of one leads to a decrease in the demand for the other

Income What would happen to your demand for milk if you lost your job? Most likely, it would fall. A lower income means that you have less to spend in total, so you would have to spend less on some – and probably most – goods. If the demand for a good falls when income falls, the good is called a **normal good**.

normal good a good for which, *ceteris paribus*, an increase in income leads to an increase in demand (and vice versa)

Not all goods are normal goods. If the demand for a good rises when income falls, the good is called an **inferior good**. An example of an inferior good might be bus rides. As your income falls, you are less likely to buy a car or take a taxi and more likely to take the bus. As income falls, therefore, demand for bus rides tends to increase.

inferior good a good for which, *ceteris paribus*, an increase in income leads to a decrease in demand (and vice versa)

Tastes A key determinant of your demand is your tastes. If you like milk, you buy more of it. Understanding the role of tastes in consumer behaviour is taking on more importance as research in the fields of psychology and neurology are applied to economics.

Population Because market demand is derived from individual demands, it follows that the more buyers there are the higher the demand is likely to be. The size of the population, therefore, is a determinant of demand. A larger population, *ceteris paribus*, will mean a higher demand for all goods and services. Changes in the way the population is structured also influences demand. Many countries have an ageing population and this leads to a change in demand. If there is an increasing proportion of the population aged 65 and over, the demand for goods and services used by the elderly, such as the demand for retirement homes, insurance policies suitable for older people, the demand for smaller cars and for health care services, etc. is likely to increase in demand as a result.

Advertising Firms advertise their products in many different ways and it is likely that if a firm embarks on an advertising campaign then the demand for that product will increase.

Expectations of Consumers Expectations about the future may affect the demand for a good or service today. For example, if it was announced that the price of milk was expected to rise next month consumers may be more willing to buy milk at today's price.

Summary The demand curve shows what happens to the quantity demanded of a good when its price varies, holding constant all the other variables that influence buyers; when one or more of these other variables changes, the demand curve shifts. Table 3.1 lists all the variables that influence how much consumers choose to buy of a good.

TABLE 3.1 Variables That Influence Buyers

This table lists the variables that affect how much consumers choose to buy of any good. Notice the special role that the price of the good plays: a change in the good's price represents a movement along the demand curve, whereas a change in one of the other variables shifts the demand curve.

| Variable | A change in this variable . . . |
|--------------------------------|---|
| Price | Is represented as a movement along the demand curve |
| Prices of other, related goods | Shifts the demand curve |
| Income | Shifts the demand curve |
| Tastes | Shifts the demand curve |
| Population | Shifts the demand curve |
| Advertising | Shifts the demand curve |
| Expectations | Shifts the demand curve |

SELF TEST Make up an example of a demand schedule for pizza and graph the demand curve. Give an example of something that would cause the demand curve to shift to the right. Give a different example of something that would cause the demand curve to shift to the left.

Finding Price and Quantity Using Algebra

In Chapter 2 we looked at linear equations and gave some examples of how demand functions can be expressed as simple linear equations.

Look at the following equation:

$$Qd = 1,700 - 3p$$

We can dissect this in a bit more detail in relation to the standard $y = a + bx$ linear equation we covered in Chapter 2.

The quantity demanded is one of the variables in the equation. In this case it is the dependent variable. The quantity demanded depends upon the price and so price is the independent variable.

The equation tells us that the quantity demanded will be 1,700 minus 3 times whatever the price is. If price is €5 then quantity demanded will be $1,700 - 3(5) = 1,685$. If price is €20 then quantity demanded will be $1,700 - 3(20) = 1,640$.

Let us take a different example. Given the demand function $p = 760 - 0.3Qd$, if price were €5 the quantity demanded (rounded up) would be:

$$\begin{aligned} 0.3Qd &= 760 - 5 \\ \frac{0.3Qd}{0.3} &= \frac{755}{0.3} \\ Qd &= 2,517 \end{aligned}$$

The same process would be used if we knew the quantity but needed to find the price. Using the equation:

$$Qd = 1,800 - 3p$$

If we knew the quantity demanded was 1,640 then substituting this into the equation would enable us to find the price:

$$\begin{aligned} Qd &= 1800 - 3p \\ Qd - 1800 &= -3p \\ \frac{Qd - 1800}{-3} &= p \\ 600 - \frac{1}{3}Qd &= p \end{aligned}$$

SELF TEST Given the equation $Qd = -8p + 2,500$, calculate the quantity demanded for prices increasing in €10 increments from €0 – €100. Plot the resulting demand curve on a piece of graph paper. At what value does the demand curve you have drawn cut the horizontal axis? What is the slope of the demand curve you have drawn? Having worked out these two values, what conclusions can you draw from comparing these values with the equation?

SUPPLY

We now turn to the other side of the market and examine the behaviour of sellers. Once again, to focus our thinking, we will continue to consider the market for milk.

The Supply Curve: The Relationship Between Price and Quantity Supplied

The **quantity supplied** of any good or service is the amount that sellers are willing and able to sell at different prices. There are many determinants of quantity supplied, but once again price plays a special role in our analysis. When the price of milk is high, selling milk is profitable, and so the sellers are willing to supply more. Sellers of milk work longer hours, buy more dairy cows and hire extra workers in order to ensure supplies to the market rise. By contrast, when the price of milk is low, the business is less profitable, and so sellers are willing to produce less milk. At a low price, some sellers may even choose to shut down, and their quantity supplied falls to zero. Because the quantity supplied rises as the price rises and falls as the price falls, we say that the quantity supplied is *positively related* to the price of the good. This relationship between price and quantity supplied is called the **law of supply**: other things equal, when the price of a good rises, the quantity supplied of the good also rises, and when the price falls, the quantity supplied falls as well.

quantity supplied the amount of a good that sellers are willing and able to sell at different prices
law of supply the claim that, *ceteris paribus*, the quantity supplied of a good rises when the price of a good rises

The table in Figure 3.4 shows the quantity, Richard, a milk producer, is willing to supply, at various prices. At a price below €0.10 per litre, Richard does not supply any milk at all. As the price rises, he is willing to supply a greater and greater quantity. This is the **supply schedule**, a table that shows the relationship between the price of a good and the quantity supplied, holding constant everything else that influences how much producers of the good want to sell.

supply schedule a table that shows the relationship between the price of a good and the quantity supplied

The graph in Figure 3.4 uses the numbers from the table to illustrate the law of supply. The curve relating price and quantity supplied is called the **supply curve**. The supply curve slopes upwards because, other things equal, a higher price means a greater quantity supplied.

supply curve a graph of the relationship between the price of a good and the quantity supplied

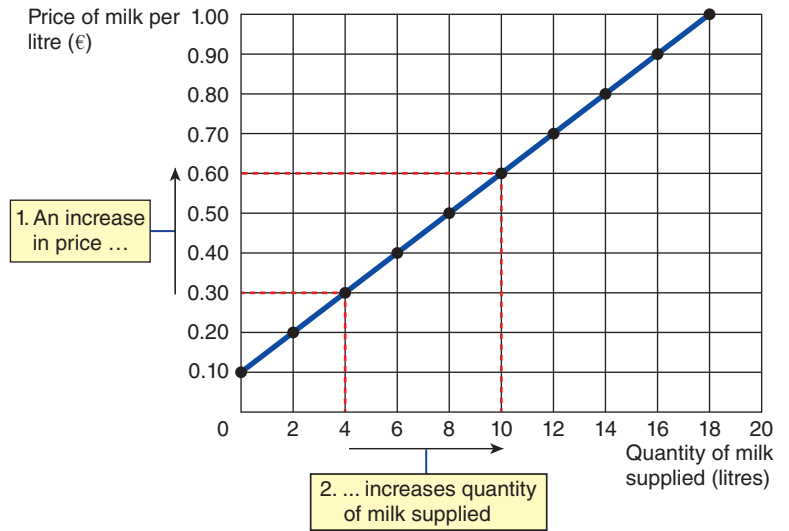
Market Supply versus Individual Supply

Just as market demand is the sum of the demands of all buyers, market supply is the sum of the supplies of all sellers. The table in Figure 3.5 shows the supply schedules for two milk producers – Richard and Megan. At any price, Richard's supply schedule tells us the quantity of milk Richard is willing to supply, and Megan's supply schedule tells us the quantity of milk Megan is willing to supply. The market supply is the sum of the two individual supplies.

The graph in Figure 3.5 shows the supply curves that correspond to the supply schedules. As with demand curves, we find the total quantity supplied at any price by adding the individual quantities found

FIGURE 3.4**Richard's Supply Schedule and Supply Curve**

The supply schedule shows the quantity supplied at each price. This supply curve, which graphs the supply schedule, shows how the quantity supplied of the good changes as its price varies. Because a higher price increases the quantity supplied, the supply curve slopes upwards.



| Price of milk per litre (€) | Quantity of milk supplied (litres per month) |
|-----------------------------|--|
| 0.00 | 0 |
| 0.10 | 0 |
| 0.20 | 2 |
| 0.30 | 4 |
| 0.40 | 6 |
| 0.50 | 8 |
| 0.60 | 10 |
| 0.70 | 12 |
| 0.80 | 14 |
| 0.90 | 16 |
| 1.00 | 18 |

on the horizontal axis of the individual supply curves. The market supply curve shows how the total quantity supplied varies as the price of the good varies.

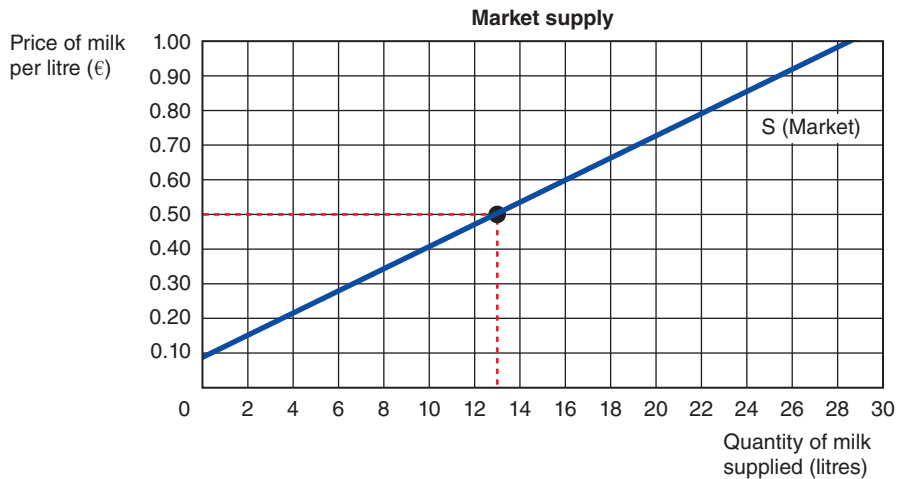
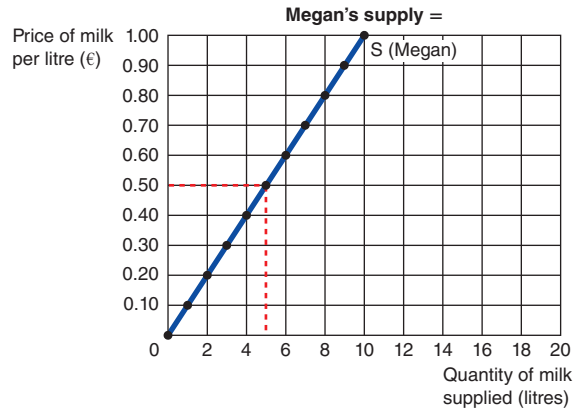
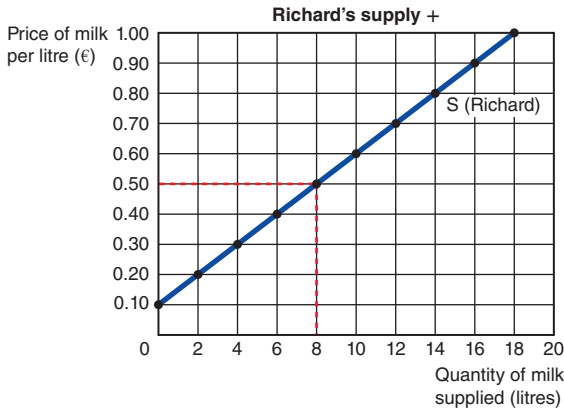
Shifts in the Supply Curve

The supply curve for milk shows how much milk producers are willing to offer for sale at any given price, holding constant all the other factors beyond price that influence producers' decisions about how much to sell. This relationship can change over time, which is represented by a shift in the supply curve. For example, suppose the price of animal feed falls. Because animal feed is an input into producing milk, the fall in the price of animal feed makes selling milk more profitable. This raises the supply of milk: at any given price, sellers are now willing to produce a larger quantity. Thus, the supply curve for milk shifts to the right.

Figure 3.6 illustrates shifts in supply. Any change that raises quantity supplied at every price, such as a fall in the price of animal feed, shifts the supply curve to the right and is called *an increase in supply*. Similarly, any change that reduces the quantity supplied at every price shifts the supply curve to the left and is called *a decrease in supply*.

FIGURE 3.5**Market Supply as the Sum of Individual Supplies**

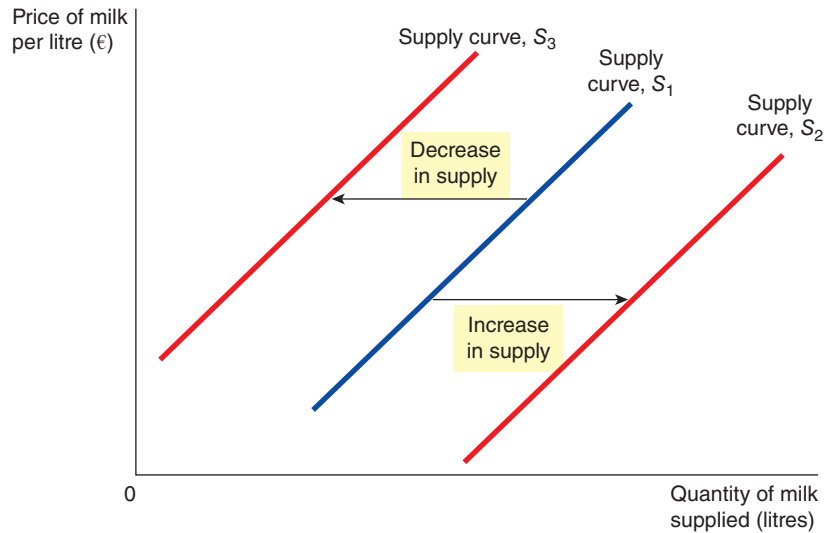
The quantity supplied in a market is the sum of the quantities supplied by all the sellers at each price. Thus, the market supply curve is found by adding horizontally the individual supply curves. At a price of €0.50, Richard is willing to supply 8 litres of milk, and Megan is willing to supply 5 litres. The quantity supplied in the market at this price is 13 litres.



| Price of milk per litre (€) | Richard + | Megan = | Market |
|-----------------------------|-----------|---------|--------|
| 0.00 | 0 | 0 | 0 |
| 0.10 | 0 | 1 | 1 |
| 0.20 | 2 | 2 | 4 |
| 0.30 | 4 | 3 | 7 |
| 0.40 | 6 | 4 | 10 |
| 0.50 | 8 | 5 | 13 |
| 0.60 | 10 | 6 | 16 |
| 0.70 | 12 | 7 | 19 |
| 0.80 | 14 | 8 | 22 |
| 0.90 | 16 | 9 | 25 |
| 1.00 | 18 | 10 | 28 |

FIGURE 3.6**Shifts in the Supply Curve**

Any change that raises the quantity that sellers wish to produce at a given price shifts the supply curve to the right. Any change that lowers the quantity that sellers wish to produce at a given price shifts the supply curve to the left.



A shift in the supply curve will be caused by one or more of the factors affecting supply **other than price**. As with demand, we can represent the factors that influence supply in a function:

$$S = f(P_n, P_{n-1}, \dots, P_{n-1}, H, N/S_i, F_1, \dots, F_m, E, S_q)$$

Where:

- P_n = Price
- P_{n-1}, \dots, P_{n-1} = Profitability of other goods in production and prices of goods in joint supply
- H = Technology
- N/S_i = Natural shocks/Social factors
- F_1, \dots, F_m = Costs of production
- E = Expectations of producers
- S_q = Number of sellers

If any of the factors affecting supply other than price changes, the supply curve will shift. The following provides a brief outline of these factors.

Profitability of Other Goods in Production and Prices of Goods in Joint Supply Firms have some flexibility in the supply of products and in some cases can switch production to other goods. For example, motor vehicle manufacturers can use production lines to change production between different models as can other manufacturers. If one product becomes more profitable then it may be that the firm switches to the more profitable product. In other cases, firms may find that products are in joint supply; an increase in the supply of lamb, for example, might also lead to an increase in the supply of wool.

Technology The technology for turning the inputs into milk is yet another determinant of supply. Advances in technology increase productivity allowing more to be produced using fewer factor inputs. As a result costs, both total and unit, may fall and supply increases. The development of fertilizers and more efficient milking parlours, for example, have increased milk yields per cow and helped reduce costs as a result. By reducing firms' costs, the advance in technology raises the supply of milk.

Natural/Social Factors There are often many natural or social factors that affect supply. These include such things as the weather affecting crops, natural disasters, pestilence and disease, changing attitudes and social expectations (for example over the production of organic food, the disposal of waste, reducing carbon emissions, ethical supply sourcing and so on) all of which can have an influence on production decisions. Some or all of these may have an influence on the cost of inputs into production.

Input Prices – the Prices of Factors of Production To produce the output of milk, sellers use various inputs including land, fertilizer, feed, silage, farm buildings, veterinary services and the labour of workers. When the price of one or more of these inputs rises, producing milk is less profitable and firms supply less milk. If input prices rise substantially, a firm might shut down and supply no milk at all. If input prices fall for some reason, then production may be more profitable and there is an incentive to supply more at each price. Thus, the supply of a good is negatively related to the price of the inputs used to make the good.

Expectations of Producers The amount of milk a firm supplies today may depend on its expectations of the future. For example, if it expects the price of milk to rise in the future, the firm might invest in more productive capacity or increase the size of the herd.

Number of Sellers If there are more sellers in the market then it makes sense that the supply would increase. Equally, if a number of dairy farms closed down then it is likely that the amount of milk supplied would also fall.

SELF TEST Make up an example of a supply schedule for pizza and graph the implied supply curve. Give an example of something that would shift this supply curve. Would a change in price shift the supply curve?

The Algebra of Supply

The principles we applied to linear demand curves can equally be applied to a linear supply curve of the type:

$$Q_s = -15 + 8p$$

The plus sign in front of the price variable in the supply equation tells us that there is a positive relationship between price and quantity supplied. Using this equation, if the price were €10 then the quantity supplied would be $-15 + 8(10) = 65$ and if price were €20 then quantity supplied would be $-15 + 8(20) = 145$.

CASE STUDY Fracking and the effect on oil prices

Developments in technology have ripple effects on many markets. In the constant search to find new sources of energy, humans have come up with many ingenious ideas and in recent years a process called fracking has caused both excitement and concern. Fracking involves injecting fluid, consisting of water and a number of chemicals, into the shale deposits underground at extremely high pressure. The pressure fractures shale rocks and releases natural gas and oil which can be 'harvested' and used for energy. The possibilities for extracting large quantities of gas and oil from fracking around the world is high but there are some who ask whether the cost in terms of the potential damage to the environment of the large amounts of water needed in the fracking process, the chemical residues left behind, the potential for contamination of drinking water supplies and pollution, not to mention concerns over the increased possibilities of earthquakes in the area around fracking wells, is worth the amount of gas and oil which can be extracted.

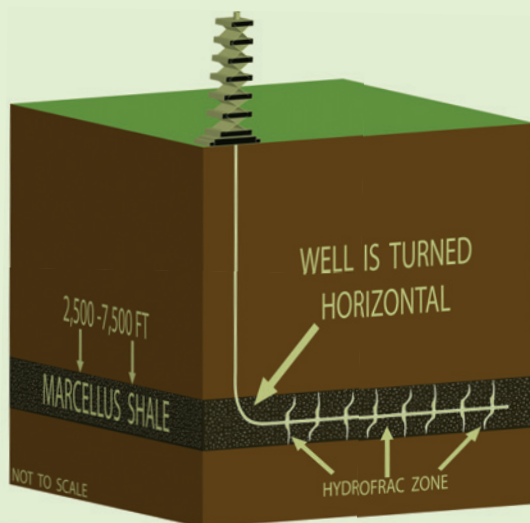
The amount of gas and oil that could be extracted is not insignificant. The chief executive officer for the US energy firm Conoco, Ryan Lance, was reported to have predicted that the United States could be self-sufficient in oil and gas by 2025. Estimates are that shale oil and gas could be around 2 million barrels per day by 2020 and could rise to 3 million barrels per day by 2035.

The Organization of Petroleum Exporting Countries (OPEC) is a group of oil-producing countries which act together to manage oil supplies. In our model of demand and supply in this chapter we have assumed that no buyer or seller can individually influence market price. Because a country like Saudi Arabia has massive reserves of oil, it can influence market price by increasing and reducing the supply of oil coming onto the market. Together, OPEC countries can and do manipulate oil supplies to try and keep the market price of oil relatively stable. Even OPEC cannot influence demand, however, and so changes in economic activity, the weather and natural disasters such as hurricanes, earthquakes and tsunamis do cause the demand for oil to change and help account for some of the volatility in market prices.

In November 2012, OPEC acknowledged that the possibilities represented by fracking might have a long-term effect on the price of oil. The predicted output of 2 million barrels per day, for example, is almost equivalent to the daily production of one of the OPEC countries, Nigeria, so the effect on oil and gas supplies in the future is going to be a factor. If fracking expands not only in North America but in the rest of the world, the effect on the supply of oil and gas will be to shift the supply curve to the right and assuming demand is relatively stable, the future oil price could fall. OPEC might choose to respond to this possibility by cutting back its own supply of oil to the market in order to stabilize the price. Because oil produced by fracking is a substitute for crude oil, OPEC expects the demand for crude oil to moderate as consumers switch to oil generated by fracking. All these factors could mean that oil prices remain more stable in the medium to long term and might also fall.

Is this good news for consumers of oil and gas? That depends on whether the concerns of environmentalists over the safety of fracking are real or not. Environmentalists would argue that the true cost

of generating shale gas and oil is not being taken into account by firms involved in the process because they are not accounting for the potential environmental damage that they say can be caused by fracking. This represents an excellent example of the working of markets and also the economic problem – are the benefits of generating more oil and gas and helping keep prices more stable, greater than the costs? Can we truly know the real costs of fracking if the companies involved dispute the claims of environmentalists?



The fracking process is controversial with opponents citing the extensive use of water in the process and the potential to cause local earthquake risks, but proponents say it is a safe and cost-effective method of generating new gas and oil supplies.

Summary The supply curve shows what happens to the quantity supplied of a good when its price varies, holding constant all the other variables that influence sellers. When one of the other variables changes, the supply curve shifts. Table 3.2 lists all the variables that influence how much producers choose to sell of a good.

TABLE 3.2 Variables That Influence Sellers

This table lists the variables that affect how much producers choose to sell of any good. Notice the special role that the price of the good plays: a change in the good's price represents a movement along the supply curve, whereas a change in one of the other variables shifts the supply curve.

| Variable | A change in this variable . . . |
|---|---|
| Price | Is represented by a movement along the supply curve |
| Profitability of goods in production and price of goods in joint supply | Shifts the supply curve |
| Technology | Shifts the supply curve |
| Natural/social factors | Shifts the supply curve |
| Input costs | Shifts the supply curve |
| Expectations of producers | Shifts the supply curve |
| Number of sellers | Shifts the supply curve |

SELF TEST Using the supply equation:

$$Q_s = 2 + 3p$$

Calculate quantity supplied if $p = 40$

Calculate price if $Q_s = 88$

SUPPLY AND DEMAND TOGETHER

Having analysed supply and demand separately, we now combine them to see how they determine the quantity of a good sold in a market and its price.

Equilibrium

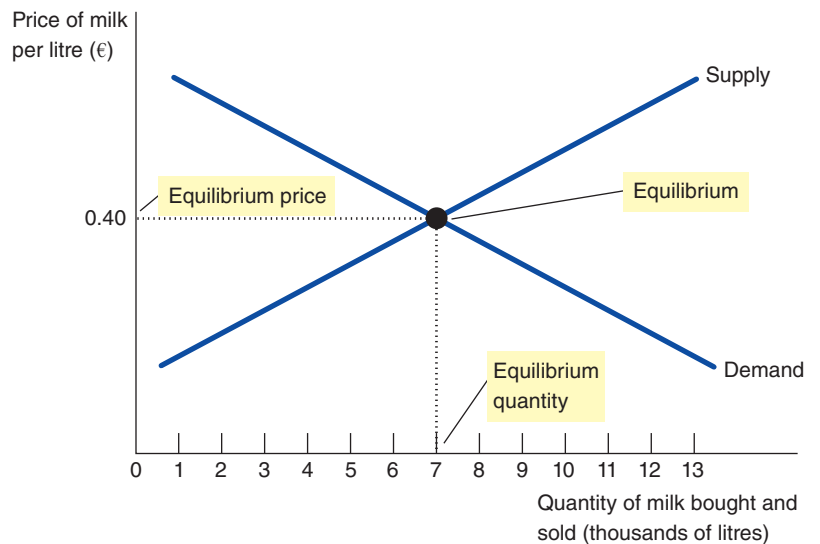
Figure 3.7 shows the market supply curve and market demand curve together. Equilibrium, remember, is defined as a state of rest, a point where there is no force acting for change. Economists refer to supply and demand as being *market forces*. In any market the relationship between supply and demand exerts force on price. If supply is greater than demand or vice versa, then there is pressure on price to change. Market equilibrium occurs when the amount consumers wish to buy at a particular price is the same as the amount sellers are willing to offer for sale at that price. The price at this intersection is called the **equilibrium or market price**, and the quantity is called the **equilibrium quantity**. In Figure 3.7 the equilibrium price is €0.40 per litre, and the equilibrium quantity is 7,000 litres of milk bought and sold.

equilibrium or market price the price where the quantity demanded is the same as the quantity supplied
equilibrium quantity the quantity bought and sold at the equilibrium price

At the equilibrium price, the quantity of the good that buyers are willing and able to buy exactly balances the quantity that sellers are willing and able to sell. The equilibrium price is sometimes called the *market-clearing price* because, at this price, everyone in the market has been satisfied: buyers have bought all they want to buy, and sellers have sold all they want to sell; there is no shortage in the market where demand is greater than supply and neither is there any surplus where supply is greater than demand.

FIGURE 3.7**The Equilibrium of Supply and Demand**

The equilibrium is found where the supply and demand curves intersect. At the equilibrium price, the quantity supplied is the same as the quantity demanded. Here the equilibrium price is €0.40 per litre of milk: at this price, sellers are willing to offer 7,000 litres of milk for sale and buyers wish to purchase 7,000 litres of milk.



The market will remain in equilibrium until something causes either a shift in the demand curve or a shift in the supply curve (or both). If one or both curves shift, at the existing equilibrium price, there will now be either a **surplus** or a **shortage**. The market mechanism takes time to adjust – sometimes it can be very quick (which tends to happen in highly organized markets like stock and commodity markets) and sometimes it is much slower to react. When the market is in disequilibrium and a shortage or surplus exists, the behaviour of buyers and sellers act as forces on price.

surplus a situation in which the quantity supplied is greater than the quantity demanded at the going market price
shortage a situation in which quantity demanded is greater than quantity supplied at the going market price

When there is a *surplus* or *excess supply* of a good, suppliers are unable to sell all they want at the going price. Sellers find stocks of milk increasing so they respond to the surplus by cutting their prices. As the price falls, some consumers are persuaded to buy more milk and so there is a movement

along the demand curve. Equally, some sellers in the market respond to the falling price by reducing the amount they are willing to offer for sale (a movement along the supply curve). Prices continue to fall until the market reaches a new equilibrium. The effect on price and the amount bought and sold depends on whether the demand curve or supply curve shifted in the first place (or whether both shifted). This is why analysis of markets is referred to as comparative statics because we are comparing one initial static equilibrium with another once market forces have worked their way through.

If the shift in demand or supply that causes the equilibrium to be disturbed creates a shortage in the market, buyers' and sellers' behaviour again 'forces' price to change to bring the market back into equilibrium. A *shortage* or situation of *excess demand* occurs where the quantity of the good demanded exceeds the quantity supplied at the going price; buyers are unable to buy all they want at that price. With too many buyers chasing too few goods, sellers can respond to the shortage by raising their prices without losing sales. As the price rises, some buyers will drop out of the market and quantity demanded falls (a movement along the demand curve). Rising prices encourage some farmers to offer more milk for sale as it is now more profitable for them to do so and the quantity supplied rises. Once again this process will continue until the market once again moves toward the equilibrium.

Thus, the activities of the many buyers and sellers 'automatically' push the market price towards the equilibrium price. Individual buyers and sellers don't consciously realize they are acting as forces for change in the market when they make their decisions but the collective act of all the many buyers and sellers does push markets towards equilibrium. This phenomenon is so pervasive that it is called the **law of supply and demand**: the price of any good adjusts to bring the quantity supplied and quantity demanded for that good into balance.

law of supply and demand the claim that the price of any good adjusts to bring the quantity supplied and the quantity demanded for that good into balance

Prices as Signals

If the government announced that there was going to be a cull of beef cattle because of some disease, if workers at major copper mines go on strike, the chances are that the price of beef and copper would rise. Markets are subject to change and changes in the factors that affect supply and demand other than price cause a shift in the curve and bring about disequilibrium. The many buyers and sellers in a market all make independent decisions which act as forces for changes in price. Why do buyers and sellers react when prices change? The reason is that price acts as a signal to both buyers and sellers.

Economists have conducted extensive research into the nature and determinants of both demand and supply. It is useful to have a little bit of background knowledge on this to help understand markets more effectively. The main function of price in a free market is to act as a signal to both buyers and sellers. For buyers, price tells them something about what they have to give up (usually an amount of money) to acquire the benefits that having the good will confer on them. These benefits are referred to as the utility or satisfaction derived from consumption. If an individual is willing to pay €10 to go and watch a movie then economists will assume that the value of the benefits gained from watching the movie is worth that amount of money to the individual. But what does this mean? How much is €10 worth? Economists would answer this question by saying that if an individual is willing to give up €10 to watch a movie then the value of the benefits gained (the utility) must be greater than the next best alternative that the €10 could have been spent on. Principles 1 and 2 of the *Ten Principles of Economics* states that people face trade-offs and that the cost of something is what you have to give up to acquire it. This is fundamental to the law of demand. At higher prices, the sacrifice being

made in terms of the value of the benefits gained from alternatives is greater and so we may be less willing to do so as a result. If the price of a ticket for the movie was €15 then it might have to be a very good movie to persuade me that giving up what else €15 could buy me is worth it. Price also acts as a signal at the margin. Most consumers will recognize the agony they have experienced over making a purchasing decision. Those ‘to die for’ pair of shoes may be absolutely perfect but at €120 they might make the buyer think twice. If they were €100 then it might be considered a ‘no brainer’. That extra €20 might make all the difference to the decision of whether to buy or not.

Economists and those in other disciplines such as psychology are increasingly investigating the complex nature of purchasing decisions that humans make. The development of magnetic resonance imaging (MRI) techniques, for example, has allowed researchers to investigate how the brain responds to different stimuli when making purchasing decisions. This increased understanding is being used by firms to persuade buyers to choose their products over rivals.

For sellers price acts as a signal in relation to the profitability of production. For most sellers, increasing the amount of a good produced will incur some additional input costs. A higher price is required in order to compensate for the additional cost and to also enable the producer to gain some reward from the risk they are taking in production. That reward is termed *profit*.

If prices are rising in a free market then this acts as a different but related signal to buyers and sellers. Rising prices to a seller means that there is a shortage and thus acts as a signal to expand production because the seller knows that s/he will be able to sell what they produce. For buyers, a rising price changes the nature of the trade-off they have to face. Rising prices act as a signal that more will have to be given up in order to acquire the good and they will have to decide whether the value of the benefits they will gain from acquiring the good is worth the extra price they have to pay and the sacrifice of the value of the benefits of the next best alternative.

For example, say the price of going to the movies increases from €10 per ticket to €15 per ticket. Some movie goers will happily pay the extra because they really enjoy a night out at the cinema but some people might start to think that €15 is a bit expensive. They might think that they could have a night out at a restaurant with friends and have a meal and a few drinks for €15 and that would represent more value to them than going to the cinema. Some of these people would, therefore, stop going to the cinema and go to a restaurant instead – the price signal to these people has changed.

What we do know is that for both buyers and sellers, there are many complex processes that occur in decision making. Whilst we do not fully understand all these processes yet, economists are constantly searching for new insights that might help them understand the workings of markets more fully. All of us go through these complex processes every time we make a purchasing decision – we may not realize it but we do! Having some appreciation of these processes is fundamental to thinking like an economist.

The Algebra of Market Equilibrium

We know that in equilibrium, demand equals supply ($D = S$). To find the market equilibrium, therefore, we set the demand and supply equations equal to each other and solve for P and Q by substitution.

Finding Market Equilibrium – The Substitution Method Take the following demand and supply equations assuming all positive quantities are greater than 20:

$$Q_d = 32 - 3p$$

$$Q_s = 20 + 4p$$

We know that in equilibrium:

$$Q_d = Q_s$$

So, equilibrium in this market will be where:

$$32 - 3p = 20 + 4p$$

This now allows us to solve for P and so find the equilibrium price:

$$32 - 3p = 20 + 4p$$

Subtract 20 from both sides and add $3p$ to both sides to get:

$$32 - 20 = 4p + 3p$$

$$12 = 7p$$

$$P = \text{€}1.71(\text{rounded to the nearest whole cent})$$

We can now substitute the equilibrium price into our two equations to find the equilibrium quantity rounded to the nearest whole number:

$$Qd = 32 - 3p$$

$$Qd = 32 - 3(1.71)$$

$$Qd = 32 - 5.13$$

$$Qd = 26.87$$

$$Qd = 27$$

As a check, follow the same process with the supply equation:

$$Qs = 20 + 4p$$

$$Qs = 20 + 4(1.71)$$

$$Qs = 20 + 4(1.71)$$

$$Qs = 26.84$$

$$Qs = 27$$

Note the figures for Qd and Qs before rounding differ slightly because we had to round the price. Now look at this example:

$$P = 3 + 0.25Qs$$

$$P = 15 - 0.75Qd$$

In this case the equations are defined in terms of price but the principle of working out equilibrium is the same as we have used above.

First, set the two equations equal to each other:

$$3 + 0.25Qs = 15 - 0.75Qd$$

Then solve for Q by adding $0.75Qd$ to both sides and then subtract 3 from both sides to get:

$$0.75Qd + 0.25Qs = 15 - 3$$

$$Q = 12$$

Substitute $Q = 12$ into one of the equations to find P .

$$P = 3 + 0.25Qs$$

$$P = 3 + 0.25(12)$$

$$P = 6$$

To check, also substitute into the demand equation:

$$P = 15 - 0.75Qd$$

$$P = 15 - 0.75(12)$$

$$P = 15 - 9$$

$$P = 6$$

Finding Market Equilibrium – The Elimination Method There is another way to find both the quantity and the price and that is through adopting the approach of solving simultaneous equations. Simultaneous equations require us to find two or more unknowns. In our case the two unknowns are price and quantity.

Look at the following two equations:

$$\begin{aligned} Q_d &= 20 - 2p \\ Q_s &= 2 + 2p \end{aligned}$$

In this case, the terms are all neatly aligned above each other so it is a relatively simple task to add the two together. Note that we are trying to find equilibrium so $Q_d = Q_s$, thus the value of Q is the same. Adding the two together we get:

$$\begin{aligned} Q_d &= 20 - 2p \\ Q_s &= 2 + 2p \\ 2Q &= 22 \\ Q &= 11 \end{aligned}$$

Notice that in the above equations we have a very convenient fact that the coefficient of p in each case is the same but with opposite signs. This makes this example very easy to eliminate p to isolate the Q value. This is not always the case, however, but it is important to remember that having two equal values with opposite signs allows us to get rid of them! We will come back to this later.

We can now use the fact that we know Q to find the equilibrium price by substituting Q into one of the equations thus:

$$\begin{aligned} Q_d &= 20 - 2p \\ 11 &= 20 - 2p \\ 2p &= 20 - 11 \\ 2p &= 9 \\ p &= 4.5 \end{aligned}$$

It is always worth checking your answer to make sure you have made no mistakes along the way so in this case we will substitute our known value of Q into the second equation to check we get the same answer ($p = 4.5$). So:

$$\begin{aligned} Q_s &= 2 + 2p \\ 11 &= 2 + 2p \\ 11 - 2 &= 2p \\ 9 &= 2p \\ p &= 4.5 \end{aligned}$$

Sometimes we may have equations where the p and Q values are both on the same side of the equation. In this case we have to use a different technique – the elimination method.

Take the following two equations:

$$\begin{aligned} -3p + 4Q &= 5 & (1) \\ 2p - 5Q &= -15 & (2) \end{aligned}$$

We have labelled these two equations (1) and (2) to allow us to keep track of what we are doing and reduce the risk of making an error.

Remember above when we noted the fact that having a nice convenient equation where the coefficient was equal but the signs were opposite, enabling us to be able to eliminate one of the values to help solve the equation for the other unknown? That is what we need to do with these two equations. We have to choose to manipulate the two equations to make either the ' p ' terms or the ' Q ' terms have the same coefficient but opposite signs. A knowledge of factors and lowest common denominators is useful here.

In this example we are going to manipulate the equations to get rid of the ' p ' terms. This allows us to isolate the ' Q ' terms and thus solve for Q and then find p .

This is how we do this:

$$-3p + 4Q = 5 \quad (1)$$

$$2p - 5Q = -15 \quad (2)$$

To eliminate p , multiply (1) by 2 and (2) by 3

$$-6p + 8Q = 10 \quad (3)$$

$$6p - 15Q = -45 \quad (4)$$

Add together (3) and (4)

$$-6p + 8Q = 10 \quad (3)$$

$$6p - 15Q = -45 \quad (4)$$

$$-7Q = -35$$

Divide both sides by -7

$$Q = 5$$

We can now substitute Q into equations (1) and (2) to find (and check) p

If $Q = 5$ then:

$$-3p + 4(5) = 5$$

$$-3p + 20 = 5$$

$$20 - 5 = 3p$$

$$15 = 3p$$

$$p = 5$$

As a check:

$$2p - 5(5) = -15$$

$$2p - 25 = -15$$

$$2p = -15 + 25$$

$$2p = 10$$

$$p = 5$$

In this example the equilibrium price is €5 and the equilibrium quantity is 5.

Three Steps to Analysing Changes in Equilibrium

So far we have seen how supply and demand together determine a market's equilibrium, which in turn determines the price of the good and the amount of the good that buyers purchase and sellers produce. Of course, the equilibrium price and quantity depend on the position of the supply and demand curves. We use comparative static analysis to look at what happens when some event shifts one of these curves and causes the equilibrium in the market to change.

To do this we proceed in three steps:

1. We decide whether the event in question shifts the supply curve, the demand curve or, in some cases, both curves.
2. We decide whether the curve shifts to the right or to the left.
3. We use the supply and demand diagram to compare the initial and the new equilibrium, which shows how the shift affects the equilibrium price and quantity.

Remember that the process by which equilibrium changes is not instantaneous – some markets will take longer to adjust to changes than others. To see how these three steps work in analysing market

changes, let's consider various events that might affect the market for milk. We begin the analysis by assuming that the market for milk is in equilibrium with the price of milk at €0.50 per litre and 13,000 litres are bought and sold and then follow our three-step approach.

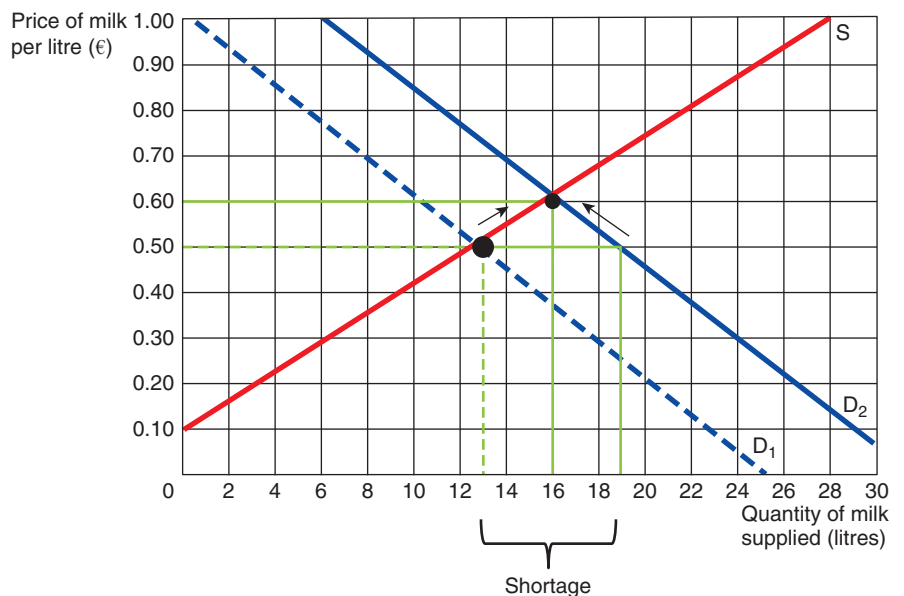
Example: A Change in Demand Suppose that one summer the weather is very hot. How does this event affect the market for milk? To answer this question, let's follow our three steps:

1. The hot weather affects the demand curve by changing peoples' taste for milk. That is, the weather changes the amount of milk that people want to buy at any given price.
2. Because hot weather makes people want to drink more milk, make refreshing milk shakes or producers of ice cream buy more milk to make ice cream, the demand curve shifts to the right. Figure 3.8 shows this increase in demand as the shift in the demand curve from D_1 to D_2 . (What you have to remember now is that demand curve D_1 does not exist anymore and so we have shown it as a dashed line.) This shift indicates that the quantity of milk demanded is higher at every price. At the existing market price of €0.50 buyers now want to buy 19,000 litres of milk but sellers are only offering 13,000 litres for sale at this price. The shift in demand has led to a shortage of milk in the market of 6,000 litres represented by the bracket.
3. The shortage encourages producers to increase the output of milk (a movement along the supply curve). There is an *increase in quantity supplied*. But, the additional production incurs extra costs and so a higher price is required to compensate sellers. As sellers increase the amount of milk offered for sale as price rises. Consumers behave differently. Some consumers who were willing to buy milk at €0.50 are not willing to pay more and so drop out of the market. As price creeps up, therefore, there is a movement along the demand curve representing these consumers who drop out of the market. The market forces of supply and demand continue to work through until a new equilibrium is reached. The new equilibrium price is now €0.60 per litre and the equilibrium quantity bought and sold is now 16,000 litres. To compare our starting and finishing positions, the hot weather which caused the shift in the demand curve has led to an increase in the price of milk and the quantity of milk bought and sold.

FIGURE 3.8

How an Increase in Demand Affects the Equilibrium

An event that raises quantity demanded at any given price shifts the demand curve to the right. The equilibrium price and the equilibrium quantity both rise. Here, an abnormally hot summer causes buyers to demand more milk. The demand curve shifts from D_1 to D_2 , which causes the equilibrium price to rise from €0.50 to €0.60 and the equilibrium quantity bought and sold to rise from 13,000 litres to 16,000 litres.



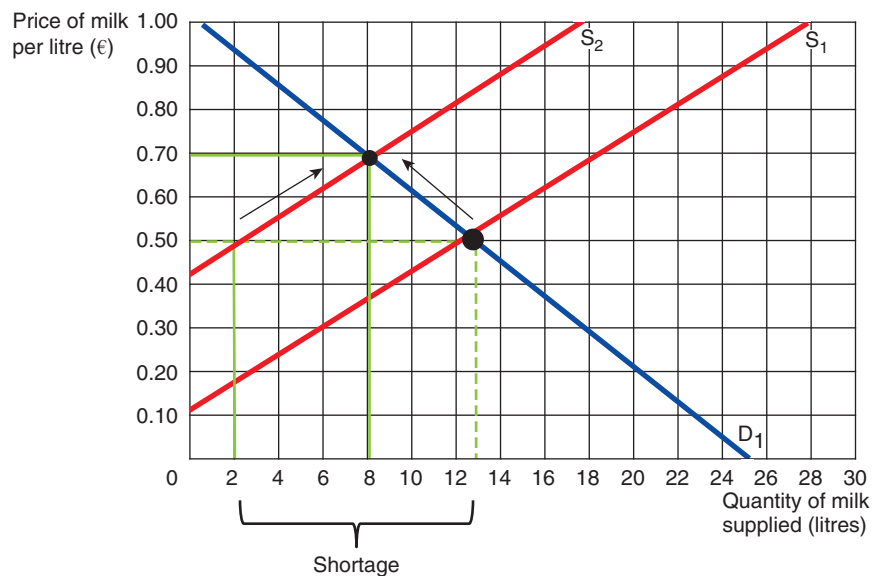
Example: A Change in Supply Suppose that, during another summer, a drought drives up the price of animal feed for dairy cattle. How does this event affect the market for milk? Once again, to answer this question, we follow our three steps:

1. The change in the price of animal feed, an input into producing milk, affects the supply curve. By raising the costs of production, it reduces the amount of milk that firms produce and sell at any given price. Some farmers may send cattle for slaughter because they cannot afford to feed them anymore and some farmers may simply decide to sell up and get out of farming altogether. The demand curve does not change because the higher cost of inputs does not directly affect the amount of milk consumers wish to buy.
2. The supply curve shifts to the left because, at every price, the total amount that farmers are willing and able to sell is reduced. Figure 3.9 illustrates this decrease in supply as a shift in the supply curve from S_1 to S_2 . At a price of €0.50 sellers are now only able to offer 2,000 litres of milk for sale but demand is still 13,000 litres. The shift in supply to the left has created a shortage in the market of 11,000 litres. Once again, the shortage will create pressure on price to rise as buyers look to purchase milk.
3. As Figure 3.9 shows, the shortage raises the equilibrium price from €0.50 to €0.70 per litre and lowers the equilibrium quantity bought and sold from 13,000 to 8,000 litres. As a result of the animal feed price increase, the price of milk rises, and the quantity of milk bought and sold falls.

FIGURE 3.9

How a Decrease in Supply Affects the Equilibrium

An event that reduces quantity supplied at any given price shifts the supply curve to the left. The equilibrium price rises, and the equilibrium quantity falls. Here, an increase in the price of animal feed (an input) causes sellers to supply less milk. The supply curve shifts from S_1 to S_2 , which causes the equilibrium price of milk to rise from €0.50 to €0.70 and the equilibrium quantity to fall from 13,000 litres to 8,000 litres.



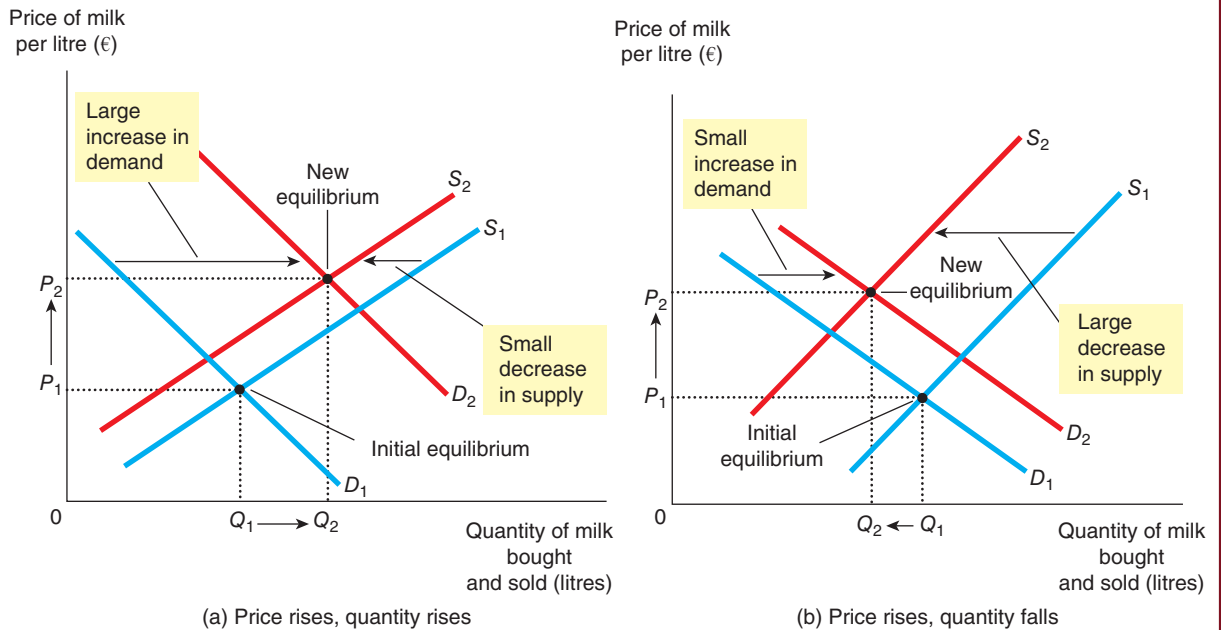
Example: A Change in Both Supply and Demand (i) Now suppose that the hot weather and the rise in animal feed occur during the same time period. To analyse this combination of events, we again follow our three steps:

1. We determine that both curves must shift. The hot weather affects the demand curve for milk because it alters the amount that consumers want to buy at any given price. At the same time, when the rise in animal feed drives up input prices, it alters the supply curve for milk because they change the amount that firms want to sell at any given price.
2. The curves shift in the same directions as they did in our previous analysis: the demand curve shifts to the right, and the supply curve shifts to the left. Figure 3.10 illustrates these shifts.
3. As Figure 3.10 shows, there are two possible outcomes that might result, depending on the relative size of the demand and supply shifts. In both cases, the equilibrium price rises. In panel (a), where demand increases substantially while supply falls just a little, the equilibrium quantity also rises.

By contrast, in panel (b), where supply falls substantially while demand rises just a little, the equilibrium quantity falls. Thus, these events certainly raise the price of milk, but their impact on the amount of milk bought and sold is ambiguous (that is, it could go either way).

FIGURE 3.10**A Shift in Both Supply and Demand (i)**

Here we observe a simultaneous increase in demand and decrease in supply. Two outcomes are possible. In panel (a), the equilibrium price rises from P_1 to P_2 , and the equilibrium quantity rises from Q_1 to Q_2 . In panel (b), the equilibrium price again rises from P_1 to P_2 , but the equilibrium quantity falls from Q_1 to Q_2 .

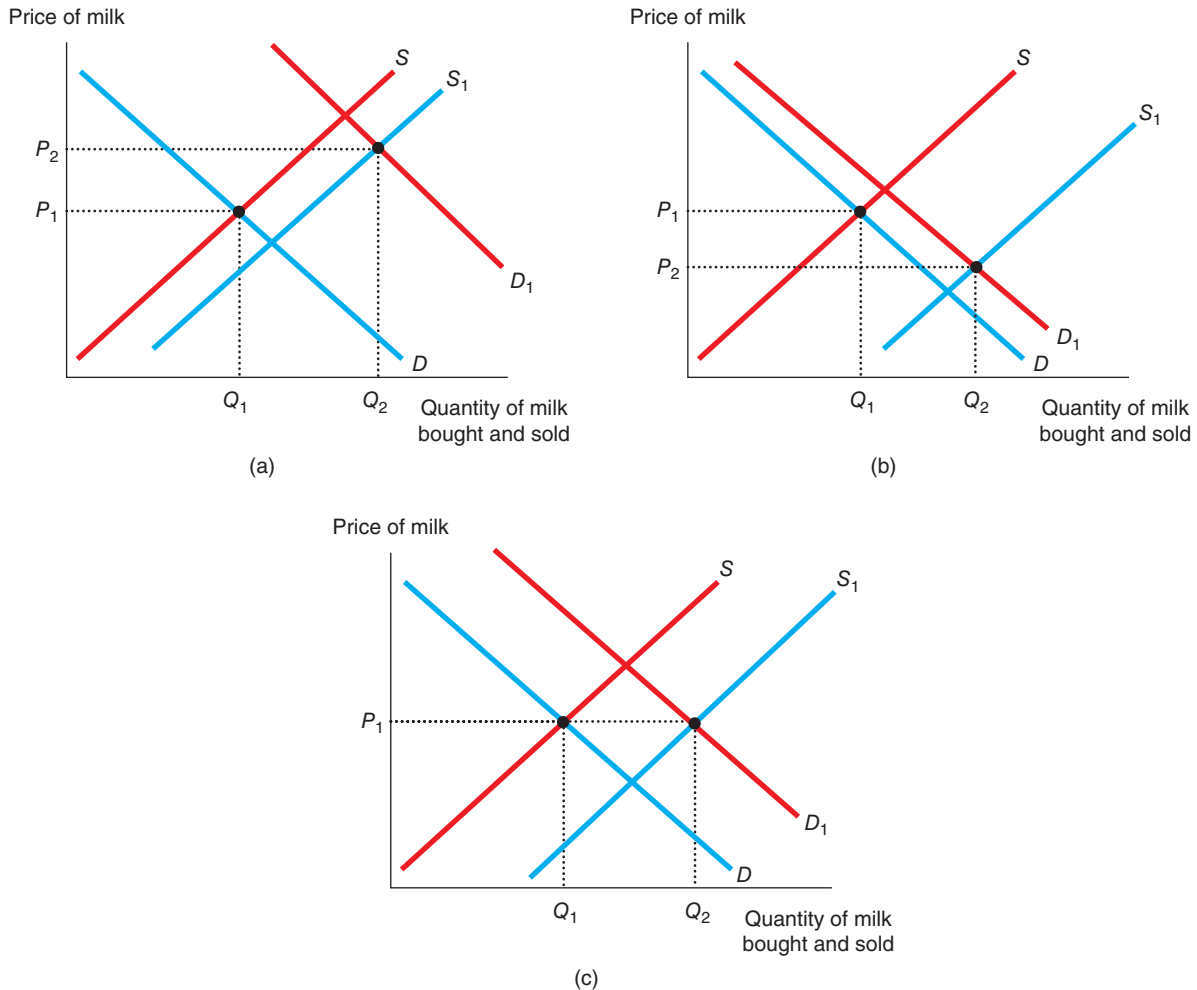


Example: A Change in Both Supply and Demand (ii) We are now going to look at a slightly different scenario but with both supply and demand changing together. Assume that forecasters have predicted a heatwave for some weeks. We know that the hot weather is likely to increase demand for milk and so the demand curve will shift to the right. However, sellers' expectations that sales of milk will increase as a result of the forecasts mean that they take steps to expand production of milk. This would lead to a shift of the supply curve to the right – more milk is now offered for sale at every price. To analyse this particular combination of events, we again follow our three steps:

1. We determine that both curves must shift. The hot weather affects the demand curve because it alters the amount of milk that consumers want to buy at any given price. At the same time, the expectations of producers alter the supply curve for milk because they change the amount that firms want to sell at any given price.
2. Both demand and supply curves shift to the right: Figure 3.11 illustrates these shifts.
3. As Figure 3.11 shows, there are three possible outcomes that might result, depending on the relative size of the demand and supply shifts. In panel (a), where demand increases substantially while supply rises just a little, the equilibrium price and quantity rises. By contrast, in panel (b), where supply rises substantially while demand rises just a little, the equilibrium price falls but the equilibrium quantity rises. In panel (c) the increases in demand and supply are identical and so equilibrium price does not change. Equilibrium quantity will increase, however. Thus, these events have different effects on the price of milk although the amount bought and sold in each case is higher. In this instance the effect on price is ambiguous.

FIGURE 3.11**A Shift in Both Supply and Demand (ii)**

Here, again, we observe a simultaneous increase in demand and supply. Here, three outcomes are possible. In panel (a) the equilibrium price rises from P_1 to P_2 and the equilibrium quantity rises from Q_1 to Q_2 . In panel (b), the equilibrium price falls from P_1 to P_2 but the equilibrium quantity rises from Q_1 to Q_2 . In panel (c), there is no change to the equilibrium price but the equilibrium quantity rises from Q_1 to Q_2 .



Summary We have just seen four examples of how to use demand and supply curves to analyse a change in equilibrium. Whenever an event shifts the demand curve, the supply curve, or perhaps both curves, you can use these tools to predict how the event will alter the amount bought and sold in equilibrium and the price at which the good is bought and sold. Table 3.3 shows the predicted outcome for any combination of shifts in the two curves. To make sure you understand how to use the tools of supply and demand, pick a few entries in this table and make sure you can explain to yourself why the table contains the prediction it does.

TABLE 3.3 What Happens to Price and Quantity When Demand or Supply Shifts?

As a test, make sure you can explain each of the entries in this table using a supply and demand diagram.

| | No change in supply | An increase in supply | A decrease in supply |
|-----------------------|---------------------|-----------------------|-----------------------|
| No change in demand | P same Q same | P down Q up | P up Q down |
| An increase in demand | P up Q up | P ambiguous Q up | P up Q ambiguous |
| A decrease in demand | P down Q down | P down Q ambiguous | P ambiguous Q down |

CONCLUSION: HOW PRICES ALLOCATE RESOURCES

This chapter has analysed demand and supply in a single market. Although our discussion has centred on the market for milk, the lessons learned here apply in most other markets as well. Whenever you go to a shop to buy something, you are contributing to the demand for that item. Whenever you look for a job, you are contributing to the supply of labour services. Because demand and supply are such pervasive economic phenomena, the model of demand and supply is a powerful tool for analysis. We shall be using this model repeatedly in the following chapters. One of the *Ten Principles of Economics* discussed in Chapter 1 is that markets are usually a good way to organize economic activity. Although it is still too early to judge whether market outcomes are good or bad, in this chapter we have begun to see how markets work. In any economic system, scarce resources have to be allocated among competing uses. Market economies harness the forces of demand and supply to serve that end. Demand and supply together determine the prices of the economy's many different goods and services; prices in turn are the signals that guide the allocation of resources.

For example, consider the allocation of property on the beach. Because the amount of this property is limited, not everyone can enjoy the luxury of living by the beach. Who gets this resource? The answer is: whoever is willing and able to pay the price. The price of seafront property adjusts until the quantity of property demanded exactly balances the quantity supplied. Thus, in market economies, prices are the mechanism for rationing scarce resources.

Similarly, prices determine who produces each good and how much is produced. For instance, consider farming. Because we need food to survive, it is crucial that some people work on farms. What determines who is a farmer and who is not? In a free society, there is no government planning agency making this decision and ensuring an adequate supply of food. Instead, the allocation of workers to farms is based on the job decisions of millions of workers. This decentralized system works well because these decisions depend on prices. The prices of food and the wages of farm workers (the price of their labour) adjust to ensure that enough people choose to be farmers.

If a person had never seen a market economy in action, the whole idea might seem preposterous. Economies are large groups of people engaged in many interdependent activities. What prevents decentralized decision making from degenerating into chaos? What coordinates the actions of the millions of people with their varying abilities and desires? What ensures that what needs to get done does in fact get done? The answer, in a word, is *prices*. If market economies are guided by an invisible hand, as Adam Smith famously suggested, then prices are the baton that the invisible hand uses to conduct the economic orchestra.

SELF TEST Analyse what happens to the market for pizza if the price of tomatoes rises.
Analyse what happens to the market for pasta if the price of potatoes falls.

IN THE NEWS



Markets in Action

We have seen how we can use supply and demand analysis to begin to understand markets. The real world has examples of markets in action every day. This article highlights an example of markets in action.

The Market for Cotton

Cotton is a commodity – a tradable product – which is used in textiles and clothing manufacture in particular, but is also used in the production of other products such as nets and filters. There are different qualities of cotton but it is reasonably homogenous. There are around 90 countries in the world that produce cotton but the United States, China and India account for around half the world's supply. There are thousands of cotton farmers and large numbers of buyers, and cotton is traded on organized

commodity markets to bring together buyers and sellers of cotton.

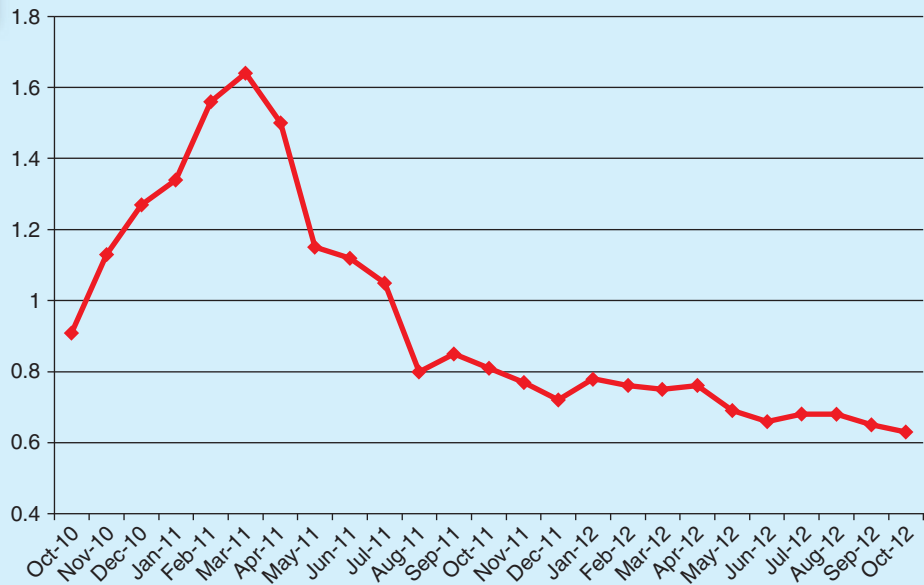
Changes in the price of cotton present an excellent way to see how markets work in real life. The price of cotton responds to the changes in demand for the commodity and to changes in supply. The supply of cotton can be dependent on factors such as the weather but also on the expectations of farmers about the future price of cotton which may influence how much land they devote to planting cotton.

What might account for these changes in prices? One of the main reasons was changes in the supply of cotton. In the years to 2010, cotton growers had faced increasing challenges making cotton farming profitable. The amount of land given over to growing cotton had declined and so supplies of cotton had been relatively low. In 2010, the weather also played a part in damaging crops and as a result there was a shortage of cotton and this accounted for the rise in price up to March 2011. How might

FIGURE 3.12

The Price of Cotton

The graph shows the price of cotton on world markets measured in euros per pound between October 2010 and October 2012. Starting at a price of around 90c per pound the price of cotton rose sharply between November 2010 and March 2011 reaching a high of €1.64 per pound before falling to around 63c per pound in October 2012.



farmers respond to the increased price? For some farmers it will now be worthwhile them planting cotton if the price they get is around €1.50–€1.60 per pound. Each farmer has to make an individual decision about whether to grow cotton but if there are large numbers of farmers across the world all making a decision to devote more land to cotton the result is going to be more cotton supplied.

A look at cotton stocks shows that between 2006 and 2009 world cotton stocks remained fairly stable at around 60 million bales. In 2010 stocks dropped to around 48 million bales. Since that time, however, stocks have risen again

and forecasters were predicting that world cotton stocks would reach around 76 million bales in 2013.

The increase in the supply of cotton has led to prices falling and those in the industry were not predicting any fundamental changes in either supply or demand over the next two years. Of course, no-one can say for sure that prices will remain stable at around 60c per pound. Some of the estimates for future stocks of cotton were based on existing acreage levels devoted to cotton and also on the assumption that nothing would happen to cause cotton harvests to change dramatically. Cotton is a crop, however, and is subject to pests

and diseases and the vagaries of the weather. No-one can tell whether the 2013 or 2014 crops might be affected by drought or flood.

Questions

- 1 To what extent would you say that the market for cotton was an example of perfect competition? Explain.**
- 2 What makes up the market for cotton?**
- 3 What signal did price send to cotton growers following the rise in cotton prices in the first quarter of 2011? How might farmers respond to this signal and what effect might this have on the future price of cotton? Explain your answer using appropriate demand and supply diagrams.**
- 4 In the article, the principle of *ceteris paribus* was implied; what do you think was assumed to be held constant? Is this a reasonable assumption to make in this instance? Explain.**
- 5 If data are collected which suggests that the amount of land in cotton-growing regions of the world devoted to cotton does not change significantly in the next five years, how useful is the model of demand and supply in predicting price changes in the cotton market in the future? Explain.**

Changes in the price of cotton presents an excellent way to see how markets work in real life. The price of cotton responds to the changes in demand for the commodity and to changes in supply.



SUMMARY

- Economists use the model of supply and demand to analyse competitive markets. In a competitive market, there are many buyers and sellers, each of whom has little or no influence on the market price.
- The demand curve shows how the quantity of a good demanded depends on the price. According to the law of demand, as the price of a good falls, the quantity demanded rises. Therefore, the demand curve slopes downwards.
- In addition to price, other determinants of how much consumers want to buy include income, the prices of substitutes and complements, tastes, expectations, the size and structure of the population and advertising. If one of these factors changes, the demand curve shifts.

- The supply curve shows how the quantity of a good supplied depends on the price. According to the law of supply, as the price of a good rises the quantity supplied rises. Therefore, the supply curve slopes upwards.
- In addition to price, other determinants of how much producers want to sell include the price and profitability of goods in production and joint supply, input prices, technology, expectations, the number of sellers and natural and social factors. If one of these factors changes, the supply curve shifts.
- The intersection of the supply and demand curves determines the market equilibrium. At the equilibrium price, the quantity demanded equals the quantity supplied.
- The behaviour of buyers and sellers naturally drives markets towards their equilibrium. When the market price is above the equilibrium price, there is a surplus of the good, which causes the market price to fall. When the market price is below the equilibrium price, there is a shortage, which causes the market price to rise.
- To analyse how any event influences a market, we use the supply and demand diagram to examine how the event affects the equilibrium price and quantity. To do this we follow three steps. First, we decide whether the event shifts the supply curve or the demand curve (or both). Second, we decide which direction the curve shifts. Third, we compare the new equilibrium with the initial equilibrium.
- In market economies, prices are the signals that guide economic decisions and thereby allocate scarce resources. For every good in the economy, the price ensures that supply and demand are in balance. The equilibrium price then determines how much of the good buyers choose to purchase and how much sellers choose to produce.

QUESTIONS FOR REVIEW

- 1 What is a competitive market? Briefly describe the types of markets other than perfectly competitive markets.
- 2 What determines the quantity of a good that buyers demand?
- 3 What are the demand schedule and the demand curve, and how are they related? Why does the demand curve slope downwards?
- 4 Does a change in consumers' tastes lead to a movement along the demand curve or a shift in the demand curve? Does a change in price lead to a movement along the demand curve or a shift in the demand curve?
- 5 Francine's income declines and, as a result, she buys more cabbage. Are cabbages an inferior or a normal good? What happens to Francine's demand curve for cabbages?
- 6 What determines the quantity of a good that sellers supply?
- 7 What are the supply schedule and the supply curve, and how are they related? Why does the supply curve slope upwards?
- 8 Does a change in producers' technology lead to a movement along the supply curve or a shift in the supply curve? Does a change in price lead to a movement along the supply curve or a shift in the supply curve?
- 9 Define the equilibrium of a market. Describe the forces that move a market toward its equilibrium.
- 10 Describe the role of prices in market economies.

PROBLEMS AND APPLICATIONS

- 1 Explain each of the following statements using supply and demand diagrams.
 - a. When there is a drought in southern Europe, the price of soft fruit rises in supermarkets throughout Europe.
 - b. When a report is published linking a product with an increased risk of cancer, the price of the product concerned tends to fall.
 - c. When conflict breaks out in the Middle East, the price of petrol in Europe rises and the price of a used Mercedes falls.
- 2 'An increase in the demand for mozzarella cheese raises the quantity of mozzarella demanded, but not the quantity supplied.' Is this statement true or false? Explain.
- 3 Technological advances have reduced the cost of producing mobile phones. How do you think this affected the market for mobile phones? For software used on mobile phones? For landlines?
- 4 Using supply and demand diagrams, show the effect of the following events on the market for sweatshirts.
 - a. A drought in Egypt damages the cotton crop.
 - b. The price of leather jackets falls.

- c. All universities require students to attend morning exercise classes in appropriate attire.
d. New knitting machines are invented.
- 5 Suppose that in the year 2005 the number of births is temporarily high. How might this baby boom affect the price of baby-sitting services in 2010 and 2020? (Hint: 5-year-olds need babysitters, whereas 15-year-olds can be babysitters.)
- 6 Think about the market for cigars.
- Are cigars substitutes or complements for cigarettes?
 - Using a supply and demand diagram, show what happens in the markets for cigars if the tax on cigarettes is increased.
 - If policymakers wanted to reduce total tobacco consumption, what policies could they combine with the cigarette tax?
- 7 The market for pizza has the following demand and supply schedules:

| Price | Quantity demanded | Quantity supplied |
|-------|-------------------|-------------------|
| €4 | 135 | 26 |
| 5 | 104 | 53 |
| 6 | 81 | 81 |
| 7 | 68 | 98 |
| 8 | 53 | 110 |
| 9 | 39 | 121 |

Graph the demand and supply curves. What is the equilibrium price and quantity in this market? If the actual price in this market were above the equilibrium price, what would drive the market towards the equilibrium? If the actual price in this market were below the equilibrium price, what would drive the market towards the equilibrium?

- 8 Because bacon and eggs are often eaten together, they are complements.
- We observe that both the equilibrium price of eggs and the equilibrium quantity of bacon have risen. What could be responsible for this pattern – a fall in the price of chicken feed or a fall in the price of pig feed? Illustrate and explain your answer.
 - Suppose instead that the equilibrium price of bacon has risen but the equilibrium quantity of eggs has fallen. What could be responsible for this pattern – a rise in the price of chicken feed or a rise in the price of pig feed? Illustrate and explain your answer.
- 9 Suppose that the price of tickets to see your local football team play at home is determined by market forces. Currently, the demand and supply schedules are as follows:

| Price | Quantity demanded | Quantity supplied |
|-------|-------------------|-------------------|
| €10 | 50,000 | 30,000 |
| 20 | 40,000 | 30,000 |
| 30 | 30,000 | 30,000 |
| 40 | 20,000 | 30,000 |
| 50 | 10,000 | 30,000 |

- Draw the demand and supply curves. What is unusual about this supply curve? Why might this be true?
 - What are the equilibrium price and quantity of tickets?
 - Your team plans to increase total capacity in its stadium by 5,000 seats next season. What admission price should it charge?
- 10 Market research has revealed the following information about the market for chocolate bars: $Q_d = 1,600 - 300p$, and the supply schedule is $Q_s = 1,400 + 700p$. Calculate the equilibrium price and quantity in the market for chocolate bars.

4

ELASTICITY AND ITS APPLICATIONS

The price a firm charges for its products is a vital part of its product positioning – what the product offering is in relation to competitors. We have seen in Chapter 3 how markets are dynamic and that price acts as a signal to both sellers and buyers; when prices change, the signal is altered and producer and consumer behaviour changes.

We know from Chapter 3 that the law of demand states that when price rises, demand falls and supply rises. What we did not discuss in the chapter was *how far* demand and supply change in response to changes in price – in other words, how sensitive supply and demand is to a change in prices. When studying how some event or policy affects a market, we discuss not only the direction of the effects but their magnitude as well. **Elasticity** is a measure of how much buyers and sellers respond to changes in market conditions and knowledge of this concept allows us to analyse supply and demand with greater precision.

elasticity a measure of the responsiveness of quantity demanded or quantity supplied to one of its determinants

THE PRICE ELASTICITY OF DEMAND

Businesses cannot directly control demand. They can seek to influence demand (and do) by utilizing a variety of strategies and tactics but ultimately the consumer decides whether to buy a product or not. One important way in which consumer behaviour can be influenced is through a firm changing the prices of its goods (many firms do have some control over the price they can charge although as we have seen, in perfectly competitive markets this is not the case as the firm is a price taker). An understanding of the price elasticity of demand is important in anticipating the likely effects of changes in price on demand.

The Price Elasticity of Demand and its Determinants

The law of demand states that a fall in the price of a good raises the quantity demanded. The **price elasticity of demand** measures how much the quantity demanded responds to a change in price. Demand for a good is said to be *elastic* or price sensitive if the quantity demanded responds substantially to changes in the price. Demand is said to be *inelastic* or price insensitive if the quantity demanded responds only slightly to changes in the price.

price elasticity of demand a measure of how much the quantity demanded of a good responds to a change in the price of that good, computed as the percentage change in quantity demanded divided by the percentage change in price

The price elasticity of demand for any good measures how willing consumers are to move away from the good as its price rises. Thus, the elasticity reflects the many economic, social and psychological forces that influence consumer tastes. Based on experience, however, we can state some general rules about what determines the price elasticity of demand.

Availability of Close Substitutes Goods with close substitutes tend to have more elastic demand because it is easier for consumers to switch from that good to others. For example, butter and margarine are easily substitutable. A relatively small increase in the price of butter, assuming the price of margarine is held fixed, causes the quantity of butter sold to fall by a relatively large amount. As a general rule, the closer the substitute the more price elastic the good is because it is easier for consumers to switch from one to the other. By contrast, because eggs are a food without a close substitute, the demand for eggs is less price elastic than the demand for butter.

Necessities versus Luxuries Necessities tend to have relatively inelastic demands, whereas luxuries have relatively elastic demands. People use gas and electricity to heat their homes and cook their food. If the price of gas and electricity rose together, people would not demand dramatically less of them. They might try and be more energy-efficient and reduce their demand a little, but they would still need hot food and warm homes. By contrast, when the price of sailing dinghies rises, the quantity of sailing dinghies sold demanded falls substantially. The reason is that most people view hot food and warm homes as necessities and a sailing dinghy as a luxury. Of course, whether a good is a necessity or a luxury depends not on the intrinsic properties of the good but on the preferences of the buyer. For an avid sailor with little concern over health issues, sailing dinghies might be a necessity with inelastic demand, and hot food and a warm place to sleep less of a necessity with a more elastic demand.

Definition of the Market The elasticity of demand in any market depends on how we draw the boundaries of the market. Narrowly defined markets tend to have more elastic demand than broadly defined markets, because it is easier to find close substitutes for narrowly defined goods. For example, food, a broad category, has a fairly inelastic demand because there are no good substitutes for food. Ice cream, a narrower category, has a more elastic demand because it is easy to substitute other desserts for ice cream. Vanilla ice cream, a very narrow category, has a very elastic demand because other flavours of ice cream are almost perfect substitutes for vanilla.

Proportion of Income Devoted to the Product Some products have a relatively high price and take a larger proportion of income than others. Buying a new suite of furniture for a lounge, for example, tends to take up a large amount of income whereas buying an ice cream might account for only a tiny proportion of income. If the price of a three-piece suite rises by 10 per cent, therefore, this is likely to have a greater effect on demand for this furniture than a similar 10 per cent increase in the price of an ice cream. The higher the proportion of income devoted to the product the greater the price elasticity is likely to be.

Time Horizon Goods tend to have more elastic demand over longer time horizons. When the price of petrol rises, the quantity of petrol demanded falls only slightly in the first few months. Over time, however, people buy more fuel-efficient cars, switch to public transport and move closer to where they work. Within several years, the quantity of petrol demanded falls more substantially. Similarly, if the price of a unit of electricity rises much above an equivalent energy unit of gas, demand may fall only slightly in the short run because many people already have electric cookers or electric heating appliances installed in their homes and cannot easily switch. If the price difference persists over several years, however, people may find it worth their while to replace their old electric heating and cooking appliances with new gas appliances and the demand for electricity will fall.

Computing the Price Elasticity of Demand

Now that we have discussed the price elasticity of demand in general terms, let's be more precise about how it is measured. Economists compute the price elasticity of demand as the percentage change in the quantity demanded divided by the percentage change in the price. That is:

$$\text{Price elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}$$

For example, suppose that a 10 per cent increase in the price of a packet of breakfast cereal causes the amount bought to fall by 20 per cent. Because the quantity demanded of a good is negatively related to its price, the percentage change in quantity will always have the opposite sign to the percentage change in price. In this example, the percentage change in price is a *positive* 10 per cent (reflecting an increase), and the percentage change in quantity demanded is a *negative* 20 per cent (reflecting a decrease). For this reason, price elasticities of demand are sometimes reported as negative numbers. In this book we follow the common practice of dropping the minus sign and reporting all price elasticities as positive numbers. (Mathematicians call this the *absolute value*.) With this convention, a larger price elasticity implies a greater responsiveness of quantity demanded to price.

Using this convention we calculate the elasticity of demand as:

$$\text{Price elasticity of demand} = \frac{20\%}{10\%} = 2$$

In this example, the price elasticity is 2, reflecting that the change in the quantity demanded is proportionately twice as large as the change in the price.

Elasticity can have a value which lies between 0 and infinity. Between 0 and 1, elasticity is said to be inelastic, that is the percentage change in quantity demanded is less than the percentage change in price. If elasticity is greater than 1 it is said to be elastic – the percentage change in quantity demanded is greater than the percentage change in price. If the percentage change in quantity demanded is the same as the percentage change in price then the elasticity is equal to 1 and is called unit or unitary elasticity.

We have and will use the term ‘relatively’ elastic or inelastic throughout our analysis. The use of this term is important. We can look at a goods, for example, both of which are classed as ‘inelastic’ but where one is more inelastic than the other. If we are comparing good *x*, which has a price elasticity of 0.2, and good *y*, which has an elasticity of 0.5, then both are price inelastic but good *y* is more price elastic by comparison. As with so much of economics, careful use of terminology is important in conveying a clear understanding.

The calculation method described above is relatively simple. You will probably find that the institution you are studying in covers a more detailed method of calculating elasticity. In this next section we will describe two such methods, the midpoint or arc elasticity of demand and point elasticity of demand. Some institutions may focus on only one of these methods in which case you can (if you wish) skip the method below which your institution does not cover.

Using the Midpoint (Arc Elasticity of Demand) Method

If you try calculating the price elasticity of demand between two points on a demand curve, you will quickly notice an annoying problem: the elasticity from point A to point B seems different from the elasticity from point B to point A. For example, consider these numbers:

Point A: Price = €4 Quantity = 120

Point B: Price = €6 Quantity = 80

The standard way to compute a percentage change is to divide the change by the initial level and multiply by 100. Going from point A to point B, the price rises by 50 per cent, and the quantity falls by 33 per cent, indicating that the price elasticity of demand is 33/50 or 0.66. By contrast, going from point B to point A, the price falls by 33 per cent, and the quantity rises by 50 per cent, indicating that the price elasticity of demand is 50/33 or 1.5.

One way to avoid this problem is to use the *midpoint method* for calculating elasticities. The midpoint method computes a percentage change by dividing the change by the midpoint (or average) of the initial and final levels. For instance, €5 is the midpoint of €4 and €6. Therefore, according to the midpoint method, a change from €4 to €6 is considered a 40 per cent rise, because $(6 - 4)/5 \times 100 = 40$. Similarly, a change from €6 to €4 is considered a 40 per cent fall.

Because the midpoint method gives the same answer regardless of the direction of change, it is often used when calculating the price elasticity of demand between two points. In our example, the midpoint between point A and point B is:

Midpoint: Price = €5, Quantity = 100

According to the midpoint method, when going from point A to point B, the price rises by 40 per cent, and the quantity falls by 40 per cent. Similarly, when going from point B to point A, the price falls by 40 per cent, and the quantity rises by 40 per cent. In both directions, the price elasticity of demand equals 1.

We can express the midpoint method with the following formula for the price elasticity of demand between two points, denoted (Q_1, P_1) and (Q_2, P_2) :

$$\text{Price elasticity of demand} = \frac{(Q_2 - Q_1)/[(Q_2 + Q_1)/2]}{(P_2 - P_1)/[(P_2 + P_1)/2]}$$

The numerator is the proportionate change in quantity computed using the midpoint method, and the denominator is the proportionate change in price computed using the midpoint method.

Using the Point Elasticity of Demand Method

Rather than measuring elasticity between two points on the demand curve, point elasticity of demand measures elasticity at a particular point on the demand curve. Let us take our general formula for price elasticity given by:

$$\text{Price elasticity of demand} = \frac{\% \Delta Qd}{\% \Delta P}$$

To calculate the percentage change in quantity demanded and the percentage change in price we use the following formulas:

$$\text{Percentage change in quantity demanded} = \frac{\Delta Qd}{Qd} \times 100$$

And

$$\text{Percentage change in price} = \frac{\Delta P}{P} \times 100$$

We can substitute these two formulas into our elasticity formula to get:

$$\text{Price elasticity of demand} = \frac{\Delta Qd}{Qd} / \frac{\Delta P}{P}$$

This can in turn be re-arranged to give:

$$\text{Price elasticity of demand} = \frac{P}{Qd} \times \frac{\Delta Qd}{\Delta P} \quad (1)$$

The slope of the demand curve is given by:

$$\text{Slope} = \frac{\Delta P}{\Delta Qd}$$

The ratio $\frac{\Delta Qd}{\Delta P}$ is the reciprocal of the slope of the demand curve. So the formula for the price elasticity of demand can also be written as:

$$\text{Price elasticity of demand} = \frac{P}{Qd} \times \frac{1}{\frac{\Delta P}{\Delta Qd}} \quad (2)$$

Using either equation 1 or equation 2 will lead to the same answer (the difference will be taking into account the negative sign, which as we have seen can be dropped when we are referring to absolute numbers).

Using calculus we get the following formula:

$$\text{Price elasticity of demand} = \frac{P}{Qd} \times \frac{d_{Qd}}{d_p}$$

This considers the change in quantity and the change in price as the ratio tends to the limit, in other words how quantity demanded responds to an infinitesimally small change in price.

The Variety of Demand Curves

Because the price elasticity of demand measures how much quantity demanded responds to changes in the price, it is closely related to the slope of the demand curve. The following heuristic (rule of thumb), is a useful guide: the flatter the demand curve that passes through a given point, the greater the price elasticity of demand. The steeper the demand curve that passes through a given point, the smaller the price elasticity of demand.

Figure 4.1 shows five cases, each of which uses the same scale on each axis. This is an important point to remember because simply looking at a graph and the shape of the curve without recognizing the scale can result in incorrect conclusions about elasticity. In the extreme case of a zero elasticity shown in panel (a), demand is *perfectly inelastic*, and the demand curve is vertical. In this case, regardless of the price, the quantity demanded stays the same. Panels (b), (c) and (d) present demand curves that are flatter and flatter, and represent greater degrees of elasticity. At the opposite extreme shown in panel (e), demand is *perfectly elastic*. This occurs as the price elasticity of demand approaches infinity and the demand curve becomes horizontal, reflecting the fact that very small changes in the price lead to huge changes in the quantity demanded.

Total Expenditure, Total Revenue and the Price Elasticity of Demand

When studying changes in demand in a market, we are interested in the amount paid by buyers of the good which will in turn represent the total revenue that sellers receive. **Total expenditure** is given by the total amount bought multiplied by the price paid. We can show total expenditure graphically, as in Figure 4.2. The height of the box under the demand curve is P and the width is Q . The area of this box, $P \times Q$, equals the total expenditure in this market. In Figure 4.2, where $P = \text{€}4$ and $Q = 100$, total expenditure is $\text{€}4 \times 100$ or $\text{€}400$.

total expenditure the amount paid by buyers, computed as the price of the good times the quantity purchased

FIGURE 4.1

The Price Elasticity of Demand

The steepness of the demand curve indicates the price elasticity of demand (assuming the scale used on the axes are the same). Note that all percentage changes are calculated using the midpoint method.

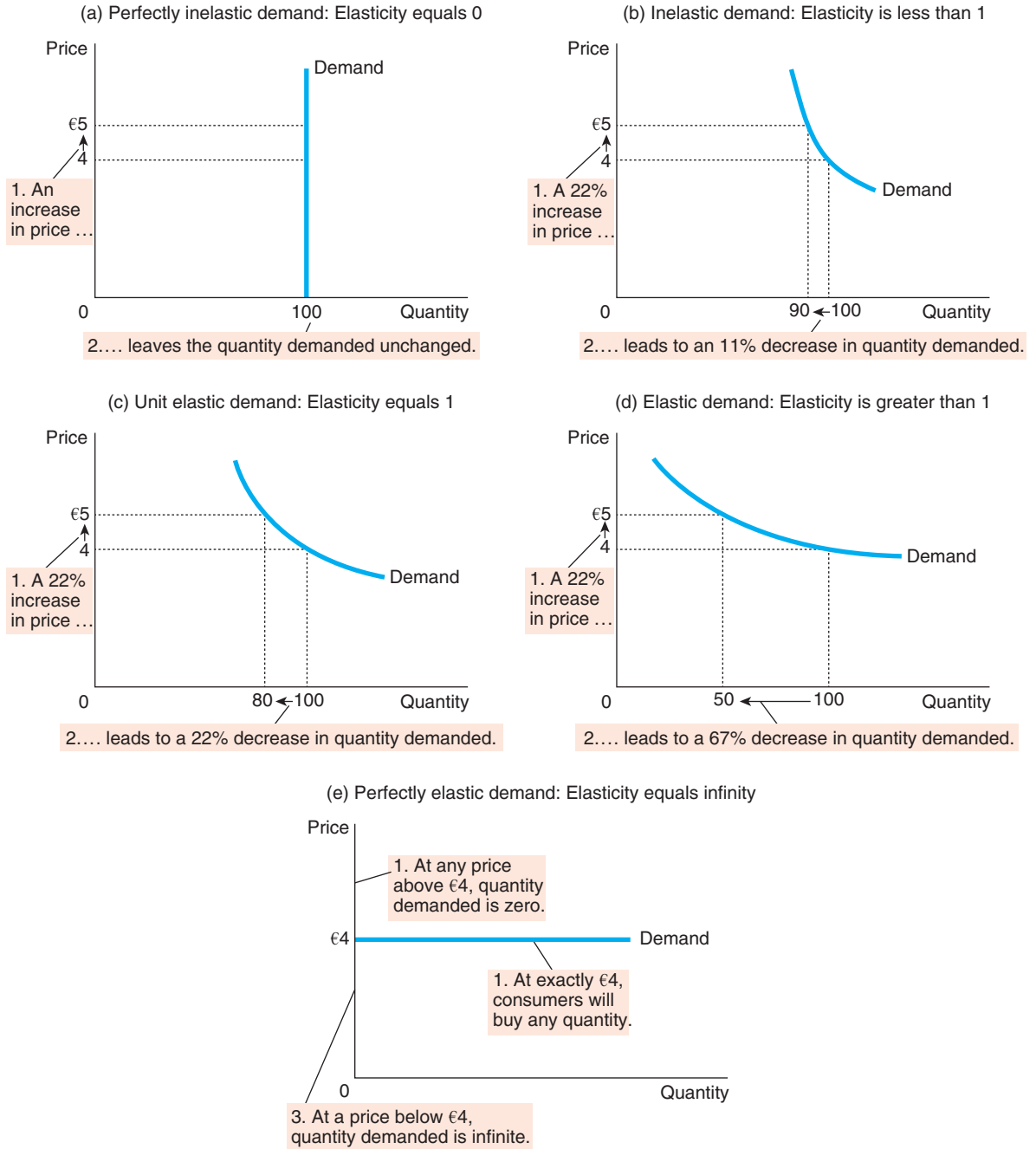
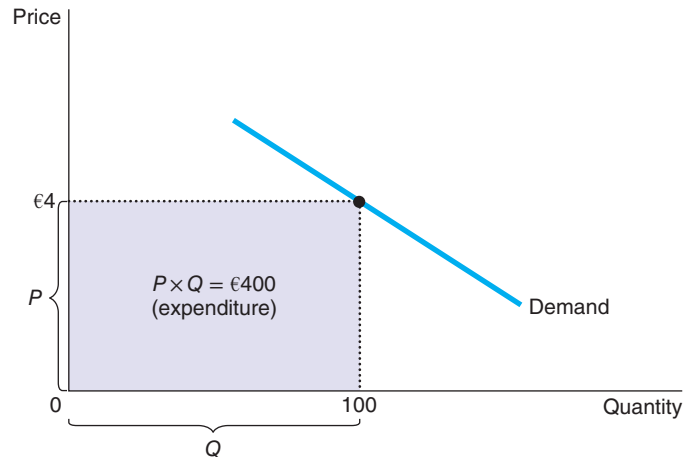


FIGURE 4.2**Total Expenditure**

The total amount paid by buyers, and received as expenditure by sellers, equals the area of the box under the demand curve, $P \times Q$. Here, at a price of €4, the quantity demanded is 100, and total expenditure is €400.



For businesses, having some understanding of the price elasticity of demand is important in decision making. If a firm is thinking of changing price how will the demand for its product react? The firm knows that there is an inverse relationship between price and demand but the effect on its revenue will be dependent on the price elasticity of demand. It is entirely possible that a firm could reduce its price and increase total revenue. Equally, a firm could raise price and find its total revenue falling. At first glance this might sound counter-intuitive but it all depends on the price elasticity of demand for the product.

If demand is inelastic, as in Figure 4.3, then an increase in the price causes an increase in total expenditure. Here an increase in price from €1 to €3 causes the quantity demanded to fall from 100 to 80, and so total expenditure rises from €100 to €240. An increase in price raises $P \times Q$ because the fall in Q is proportionately smaller than the rise in P .

We obtain the opposite result if demand is elastic: an increase in the price causes a decrease in total expenditure. In Figure 4.4, for instance, when the price rises from €4 to €5, the quantity demanded falls from 50 to 20, and so total expenditure falls from €200 to €100. Because demand is elastic, the reduction in the quantity demanded is so great that it more than offsets the increase in the price. That is, an increase in price reduces $P \times Q$ because the fall in Q is proportionately greater than the rise in P .

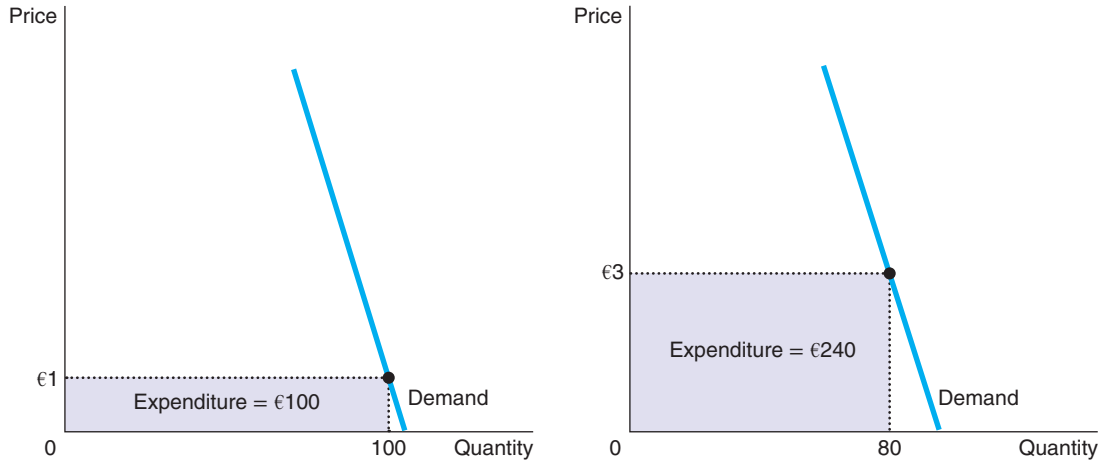
Although the examples in these two figures are extreme, they illustrate a general rule:

- When demand is inelastic (a price elasticity less than 1), price and total expenditure move in the same direction.
- When demand is elastic (a price elasticity greater than 1), price and total expenditure move in opposite directions.
- If demand is unit elastic (a price elasticity exactly equal to 1), total expenditure remains constant when the price changes.

Economists have attempted to place some estimates on the price elasticity of demand for certain goods – these do vary depending on which source you are looking at and the methods used by the researcher but Table 4.1 provides a summary of some of these estimates.

FIGURE 4.3**How Total Expenditure Changes When Price Changes: Inelastic Demand**

With an inelastic demand curve, an increase in the price leads to a decrease in quantity demanded that is proportionately smaller. Therefore, total expenditure (the product of price and quantity) increases. Here, an increase in the price from €1 to €3 causes the quantity demanded to fall from 100 to 80, and total expenditure rises from €100 to €240.

**FIGURE 4.4****How Total Expenditure Changes When Price Changes: Elastic Demand**

With an elastic demand curve, an increase in the price leads to a decrease in quantity demanded that is proportionately larger. Therefore, total expenditure (the product of price and quantity) decreases. Here, an increase in the price from €4 to €5 causes the quantity demanded to fall from 50 to 20, so total expenditure falls from €200 to €100.



TABLE 4.1**Estimates of the Price Elasticity of Demand for a Selection of Goods**

| Product | Price Elasticity of Demand |
|-------------------------|----------------------------|
| Tobacco | 0.4 |
| Milk | 0.3 |
| Wine | 0.6 |
| Shoes | 0.7 |
| Cars | 1.9 |
| Particular brand of car | 4.0 |
| Movies | 0.9 |
| Entertainment | 1.4 |
| Furniture | 3.04 |
| Fuel | 0.4 |
| Bread | 0.25 |

Elasticity and Total Expenditure along a Linear Demand Curve

Demand curves can be linear (straight) or can be curvilinear (curved). The elasticity at any point along a demand curve will depend on the shape of the demand curve. A linear demand curve has a constant slope. Recall that slope is defined as 'rise-over-run', which here is the ratio of the change in price ('rise') to the change in quantity ('run'). The slope of the demand curve in Figure 4.5 is constant because each €1 increase in price causes the same 2-unit decrease in the quantity demanded.

FIGURE 4.5**Elasticity of a Linear Demand Curve**

The slope of a linear demand curve is constant, but its elasticity is not. The demand schedule in the table was used to calculate the price elasticity of demand by the midpoint method. At points with a low price and high quantity, the demand curve is inelastic. At points with a high price and low quantity, the demand curve is elastic.

| Price | Quantity | Total revenue (Price × Quantity) | Per cent change in price | Per cent change in quantity | Price elasticity | Quantity description |
|-------|----------|-------------------------------------|-----------------------------|--------------------------------|---------------------|-------------------------|
| €7 | 0 | €0 | 15 | 200 | 13.0 | Elastic |
| 6 | 2 | 12 | 18 | 67 | 3.7 | Elastic |
| 5 | 4 | 20 | 22 | 40 | 1.8 | Elastic |
| 4 | 6 | 24 | 29 | 29 | 1.0 | Unit elastic |
| 3 | 8 | 24 | 40 | 22 | 0.6 | Inelastic |
| 2 | 10 | 20 | 67 | 18 | 0.3 | Inelastic |
| 1 | 12 | 12 | 200 | 15 | 0.1 | Inelastic |
| 0 | 14 | 0 | | | | |



Even though the slope of a linear demand curve is constant, the elasticity is not. The reason is that the slope is the ratio of *changes* in the two variables, whereas the elasticity is the ratio of *percentage changes* in the two variables. You can see this by looking at the table in Figure 4.5, which shows the demand schedule for the linear demand curve in the graph. The table uses the midpoint method to calculate the price elasticity of demand. At points with a low price and high quantity, the demand curve is inelastic. At points with a high price and low quantity, the demand curve is elastic.

The table also presents total expenditure at each point on the demand curve. These numbers illustrate the relationship between total expenditure and elasticity. When the price is €1, for instance, demand is inelastic and a price increase to €2 raises total expenditure. When the price is €5, demand is elastic, and a price increase to €6 reduces total expenditure. Between €3 and €4, demand is exactly unit elastic and total expenditure is the same at these two prices.

OTHER DEMAND ELASTICITIES

In addition to the price elasticity of demand, economists also use other elasticities to describe the behaviour of buyers in a market.

The Income Elasticity of Demand

The **income elasticity of demand** measures how the quantity demanded changes as consumer income changes. It is calculated as the percentage change in quantity demanded divided by the percentage change in income. That is:

$$\text{Income elasticity of demand} = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in income}}$$

income elasticity of demand a measure of how much the quantity demanded of a good responds to a change in consumers' income, computed as the percentage change in quantity demanded divided by the percentage change in income

Most goods are *normal goods*: higher income raises quantity demanded. Because quantity demanded and income change in the same direction, normal goods have positive income elasticities. A few goods, such as bus rides, are *inferior goods*: higher income lowers the quantity demanded. Because quantity demanded and income move in opposite directions, inferior goods have negative income elasticities.

Even among normal goods, income elasticities vary substantially in size. Necessities, such as food and clothing, tend to have small income elasticities because consumers, regardless of how low their incomes, choose to buy some of these goods. Luxuries, such as caviar and diamonds, tend to have high income elasticities because consumers feel that they can do without these goods altogether if their income is too low.

The Cross-Price Elasticity of Demand

The **cross-price elasticity of demand** measures how the quantity demanded of one good changes as the price of another good changes. It is calculated as the percentage change in quantity demanded of good 1 divided by the percentage change in the price of good 2. That is:

$$\text{Cross-price elasticity of demand} = \frac{\text{Percentage change in quantity demanded of good 1}}{\text{Percentage change in the price of good 2}}$$

cross-price elasticity of demand a measure of how much the quantity demanded of one good responds to a change in the price of another good, computed as the percentage change in quantity demanded of the first good divided by the percentage change in the price of the second good

CASE STUDY Demand for Seafood in China

At the time of writing there is still a significant proportion of the Chinese population who live in abject poverty. However, rapid economic growth in China in the past 20 years has meant that there is now a population group who are becoming increasingly wealthy in that their incomes are rising. When incomes rise, people are able to purchase more goods and services, and this can have an effect on some industries more than others.

The effect will be dependent on the income elasticity of demand. Rabobank, a cooperative financial services provider, estimates that the income elasticity of demand for fish in China is much higher than for meat and as incomes rise this has implications for those in the fishing industry. There will be firms in China involved in catching and processing fish, of course, but there may also be opportunities for firms in other countries to boost exports of fish to China. The ability of China to satisfy the growing demand for fish is likely to be constrained by the amount of resources available to satisfy the growing demand – in other words, the price elasticity of supply is relatively inelastic in the short run. This leaves gaps in the market which can be satisfied by exporters. For example, Rabobank reports that European exports of Atlantic salmon have risen from extremely low levels to in excess of 12,000 tonnes since 2000 and the total seafood export market to China could be worth around €15.7 billion by 2020. If income elasticity of demand was much lower then the opportunities for fish exporters in Europe would be much less.



Fish farms like this in China may help to meet some of the rising demand for fish in the country but the price elasticity of supply of fish in the short run is likely to be inelastic.

Whether the cross-price elasticity is a positive or negative number depends on whether the two goods are substitutes or complements. Substitutes are goods that are typically used in place of one another, such as beef steak and Wiener schnitzel. An increase in the price of beef steak induces people to eat Wiener schnitzel instead. Because the price of beef steak and the quantity of Wiener schnitzel demanded move in the same direction, the cross-price elasticity is positive. Conversely, complements are goods that are typically used together, such as computers and software. In this case, the cross-price elasticity is negative, indicating that an increase in the price of computers reduces the quantity of software demanded. As with price elasticity of demand, cross-price elasticity may increase over time: a change in the price of electricity will have little effect on demand for gas in the short run but much stronger effects over several years.

SELF TEST Define the price elasticity of demand. • Explain the relationship between total expenditure and the price elasticity of demand.

PRICE ELASTICITY OF SUPPLY

When we introduced supply in Chapter 3, we noted that producers of a good offer to sell more of it when the price of the good rises, when their input prices fall or when their technology improves. To turn from qualitative to quantitative statements about quantity supplied, we once again turn to the concept of elasticity.

The Price Elasticity of Supply and its Determinants

The law of supply states that higher prices raise the quantity supplied. The **price elasticity of supply** measures how much the quantity supplied responds to changes in the price. Supply of a good is said to be *elastic* (or price sensitive) if the quantity supplied responds substantially to changes in the price. Supply is said to be *inelastic* (or price insensitive) if the quantity supplied responds only slightly to changes in the price.

price elasticity of supply a measure of how much the quantity supplied of a good responds to a change in the price of that good, computed as the percentage change in quantity supplied divided by the percentage change in price

The price elasticity of supply depends on the flexibility of sellers to change the amount of the good they produce. For example, seafront property has an inelastic supply because it is almost impossible to produce more of it quickly – supply is not very sensitive to changes in price. By contrast, manufactured goods, such as books, cars and television sets, have relatively elastic supplies because the firms that produce them can run their factories longer in response to a higher price – supply is sensitive to changes in price.

Elasticity can take any value greater than or equal to 0. The closer to 0 the more inelastic, and the closer to infinity the more elastic.

The Determinants of Price Elasticity of Supply

The Time Period In most markets, a key determinant of the price elasticity of supply is the time period being considered. Supply is usually more elastic in the long run than in the short run. We can further distinguish between the short run and the very short run. Over very short periods of time, firms may find it impossible to respond to a change in price by changing output. In the short run firms cannot easily change the size of their factories or productive capacity to make more or less of a good but may have some flexibility. For example, it might take a month to employ new labour but after that time some increase in output can be accommodated. Overall, in the short run, the quantity supplied is not very responsive to the price. By contrast, over longer periods, firms can build new factories or close old ones, hire new staff, and buy in more capital and equipment. In addition, new firms can enter a market and old firms can shut down. Thus, in the long run, the quantity supplied can respond substantially to price changes.

Productive Capacity Most businesses, in the short run, will have a finite capacity – an upper limit to the amount that they can produce at any one time determined by the amount of factor inputs they possess. How far they are using this capacity depends, in turn, on the state of the economy. In periods of strong economic growth, firms may be operating at or near full capacity. If demand is rising for the product they produce and prices are rising, it may be difficult for the firm to expand output to meet this new demand and so supply may be inelastic.

When the economy is growing slowly or is contracting, some firms may find they have to cut back output and may only be operating at 60 per cent of full capacity. In this situation, if demand later increased and prices started to rise, it may be much easier for the firm to expand output relatively quickly and so supply would be more elastic.

The Size of the Firm/Industry It is possible that as a general rule, supply may be more elastic in smaller firms or industries than in larger ones. For example, consider a small independent furniture manufacturer. Demand for its products may rise and in response the firm may be able to buy in raw materials (wood, for example),

to meet this increase in demand. Whilst the firm will incur a cost in buying in this timber, it is unlikely that the unit cost for the material will increase. Compare this to a situation where a steel manufacturer increases its purchase of raw materials (iron ore, for example). Buying large quantities of iron ore on global commodity markets can drive up unit price and, by association, unit costs.

A study of the coffee industry in Papua New Guinea¹ found that the price elasticity of supply of smallholders was around 0.23 whereas that for large estates was much more inelastic at 0.04. The study suggested that these figures compared well to estimates derived in previous studies.

The response of supply to changes in price in large firms/industries, therefore, may be less elastic than in smaller firms/industries. This is also related to the number of firms in the industry – the more firms there are in the industry the easier it is to increase supply, *ceteris paribus*.

The Mobility of Factors of Production Consider a farmer whose land is currently devoted to producing wheat. A sharp rise in the price of rape seed might encourage the farmer to switch use of land from wheat to rape seed relatively easily. The mobility of the factor of production land, in this case, is relatively high and so supply of rape seed may be relatively elastic.

A number of multinational firms that have plants in different parts of the world now build each plant to be identical. What this means is that if there is disruption to one plant the firm can more easily transfer operations to another plant elsewhere and continue production ‘seamlessly’. Car manufacturers provide another example of this interchangeability of parts and operations. The chassis, for example, may be identical across a range of branded car models. This is the case with some Audi, Volkswagen, Seat and Skoda models. This means that the supply may be more elastic as a result.

Compare this to the supply of highly skilled oncology consultants. An increase in the wages of oncology consultants (suggesting a shortage exists) will not mean that a renal consultant or other doctors can suddenly switch to take advantage of the higher wages and increase the supply of oncology consultants. In this example, the mobility of labour to switch between different uses is limited and so the supply of these specialist consultants is likely to be relatively inelastic.

Ease of Storing Stock/Inventory In some firms, stocks can be built up to enable the firm to respond more flexibly to changes in prices. In industries where inventory build-up is relatively easy and cheap, the price elasticity of supply is more elastic than in industries where it is much harder to do this. Consider the fresh fruit industry, for example. Storing fresh fruit is not easy because it is perishable and so the price elasticity of supply in this industry may be more inelastic.

Table 4.2 presents some estimates of the price elasticity of supply.

TABLE 4.2**Estimates of Price Elasticity of Supply**

| Good | Price Elasticity of Supply Estimate |
|---|--|
| Public transport in Sweden | 0.44 to 0.64 |
| Labour in South Africa | 0.35 to 1.75 |
| Beef | |
| • Zimbabwe | 2.0 |
| • Brazil | 0.11 to 0.56 |
| • Argentina | 0.67 to 0.96 |
| Corn (short run in US) | 0.96 |
| Housing, long-run in selected US cities | Dallas: 38.6 San Francisco: 2.4 New Orleans: 0.9 St. Louis: 8.1 |
| Uranium | 2.3 to 3.3 |
| Recycled aluminium | 0.5 |
| Oysters | 1.64 to 2.00 |
| Retail store space | 3.2 |
| Natural gas (short-run) | 0.5 |

¹<http://www.une.edu.au/bepp/working-papers/ag-res-econ/occasional-papers/agecop3.pdf>

Computing the Price Elasticity of Supply

Computing the price elasticity of supply is similar to the process we adopted for calculating the price elasticity of demand and the two methods of calculating price elasticity of demand, the midpoint or arc method and the point elasticity method also applies to supply.

The price elasticity of supply is the percentage change in the quantity supplied divided by the percentage change in the price. That is:

$$\text{Price elasticity of supply} = \frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}}$$

For example, suppose that a 10 per cent increase in the price of bicycles causes the amount of bicycles supplied to the market to rise by 15 per cent. We calculate the elasticity of supply as:

$$\begin{aligned}\text{Price elasticity of supply} &= \frac{15}{10} \\ \text{Price elasticity of supply} &= 1.5\end{aligned}$$

In this example, the elasticity of 1.5 reflects the fact that the quantity supplied moves proportionately one and a half times as much as the price.

The Midpoint (Arc) Method of Calculating Percentage Changes and Elasticities

As with the price elasticity of demand, we use the midpoint method with the following formula for the price elasticity of supply between two points, denoted (Q_1, P_1) and (Q_2, P_2) :

$$\text{Price elasticity of supply} = \frac{(Q_2 - Q_1)/[(Q_2 + Q_1)/2]}{(P_2 - P_1)/[(P_2 + P_1)/2]}$$

The numerator is the percentage change in quantity supplied computed using the midpoint method, and the denominator is the percentage change in price computed using the midpoint method.

For example, suppose that an increase in the price of milk from €2.85 to €3.15 a litre raises the amount that dairy farmers produce from 90,000 to 110,000 litres per month. Using the midpoint method, we calculate the percentage change in price as:

$$\text{Percentage change in price} = (3.15 - 2.85)/3.00 \times 100 = 10 \text{ per cent}$$

Similarly, we calculate the percentage change in quantity supplied as:

$$\text{Percentage change in quantity supplied} = (110,000 - 90,000)/100,000 \times 100 = 20 \text{ per cent}$$

In this case, the price elasticity of supply is:

$$\text{Price elasticity of supply} = \frac{20\%}{10\%} = 2$$

In this example, the elasticity of 2 reflects the fact that the quantity supplied moves proportionately twice as much as the price.

Point Elasticity of Supply Method

As with point elasticity of demand, point elasticity of supply measures elasticity at a particular point on the supply curve. Exactly the same principles apply as with point elasticity of demand so the formula for point elasticity of supply is given by:

$$\text{Price elasticity of supply} = \frac{\% \Delta Q_s}{\% \Delta P}$$

Let us look at an example:

Given the supply function, $Q_s = -2 + 5p$, if price changes by 1, quantity supplied will change by 5.

The ratio, $\frac{\Delta Q_s}{\Delta P}$ (the reciprocal of the slope, remember) in this case will be $\frac{5}{1} = 5$.

Using calculus we get the following formula:

$$\text{Price elasticity of demand} = \frac{P}{Q_s} \times \frac{d_{Q_s}}{d_p}$$

This considers the change in quantity supplied and the change in price as the ratio tends to the limit, in other words how quantity supplied responds to an infinitesimally small change in price.

The Variety of Supply Curves

Because the price elasticity of supply measures the responsiveness of quantity supplied to the price, it is reflected in the appearance of the supply curve (again, assuming we are using similar scales on the axes of diagrams being used). Figure 4.6 shows five cases. In the extreme case of a zero elasticity, as shown in panel (a), supply is *perfectly inelastic* and the supply curve is vertical. In this case, the quantity supplied is the same regardless of the price. In panels (b), (c) and (d) the supply curves are increasingly flatter associated with increasing price elasticity, which shows that the quantity supplied responds more to changes in the price. At the opposite extreme, shown in panel (e), supply is *perfectly elastic*. This occurs as the price elasticity of supply approaches infinity and the supply curve becomes horizontal, meaning that very small changes in the price lead to very large changes in the quantity supplied.

In some markets, the elasticity of supply is not constant but varies over the supply curve. Figure 4.7 shows a typical case for an industry in which firms have factories with a limited capacity for production. For low levels of quantity supplied, the elasticity of supply is high, indicating that firms respond substantially to changes in the price. In this region, firms have capacity for production that is not being used, such as buildings and machinery sitting idle for all or part of the day. Small increases in price make it profitable for firms to begin using this idle capacity. As the quantity supplied rises, firms begin to reach capacity. Once capacity is fully used, increasing production further requires the construction of new factories. To induce firms to incur this extra expense, the price must rise substantially, so supply becomes less elastic.

Figure 4.7 presents a numerical example of this phenomenon. In each case below we have used the midpoint method and the numbers have been rounded for convenience. When the price rises from €3 to €4 (a 29 per cent increase, according to the midpoint method), the quantity supplied rises from 100 to 200 (a 67 per cent increase). Because quantity supplied moves proportionately more than the price, the supply curve has elasticity greater than 1. By contrast, when the price rises from €12 to €15 (a 22 per cent increase), the quantity supplied rises from 500 to 525 (a 5 per cent increase). In this case, quantity supplied moves proportionately less than the price, so the elasticity is less than 1.

SELF TEST Define the price elasticity of supply. • Explain why the price elasticity of supply might be different in the long run from in the short run.

FIGURE 4.6

The Price Elasticity of Supply

The price elasticity of supply determines whether the supply curve is steep or flat (assuming that the scale used for the axes is the same). Note that all percentage changes are calculated using the midpoint method.

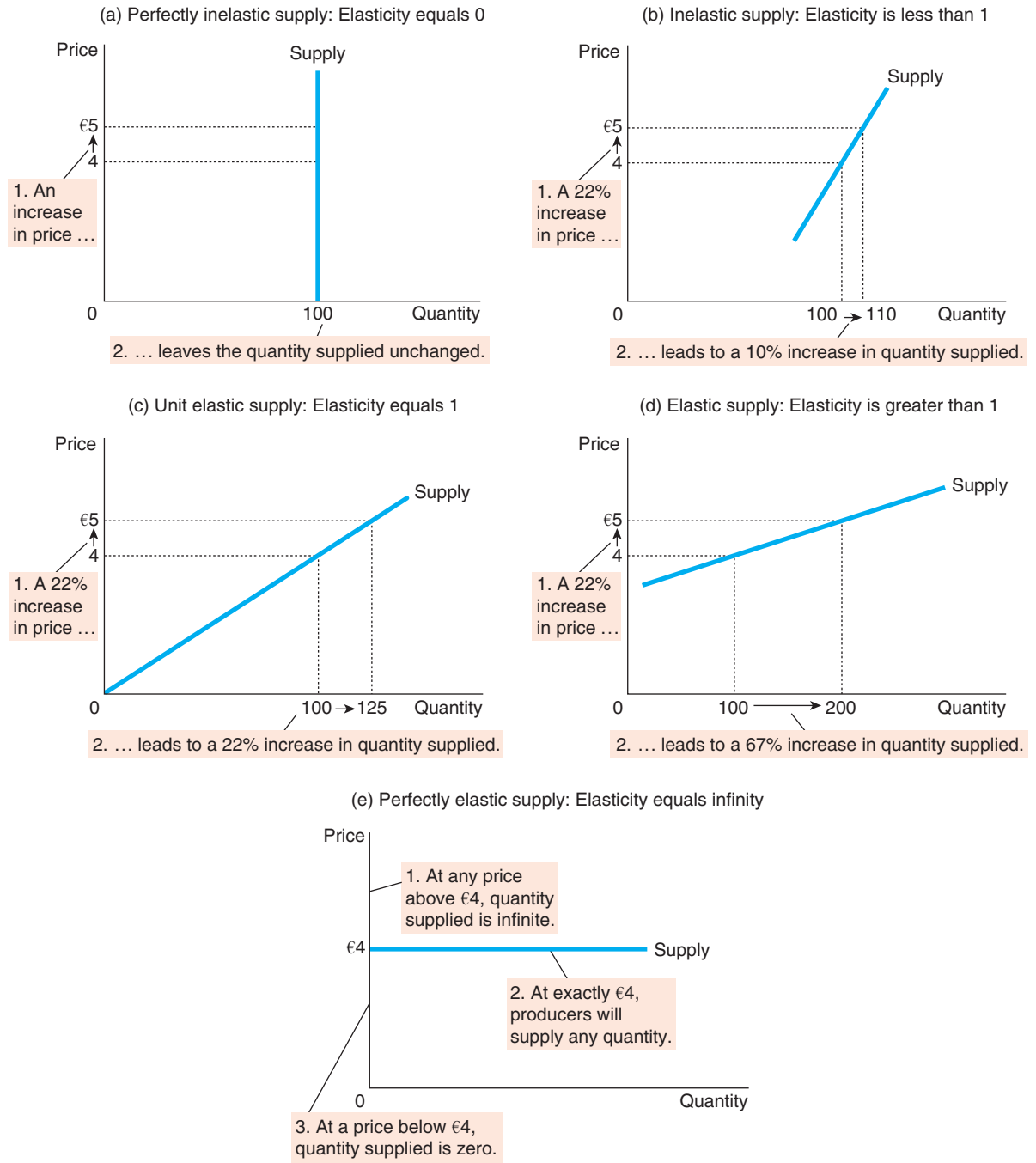
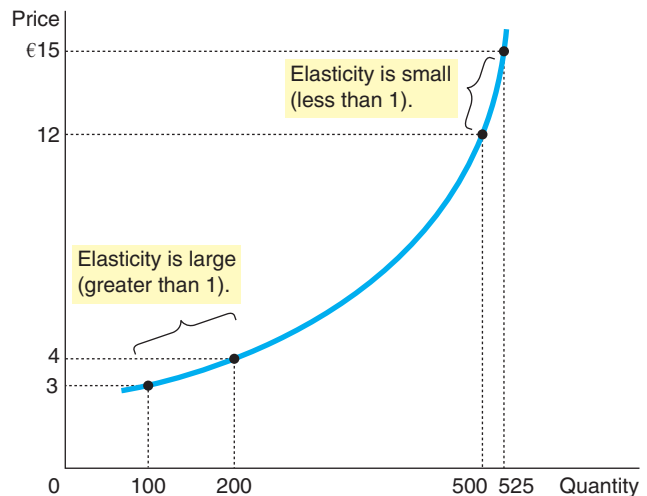


FIGURE 4.7**How the Price Elasticity of Supply Can Vary**

Because firms often have a maximum capacity for production, the elasticity of supply may be very high at low levels of quantity supplied and very low at high levels of quantity supplied. Here, an increase in price from €3 to €4 increases the quantity supplied from 100 to 200. Because the increase in quantity supplied of 67 per cent (computed using the midpoint method) is larger than the increase in price of 29 per cent, the supply curve is elastic in this range. By contrast, when the price rises from €12 to €15, the quantity supplied rises only from 500 to 525. Because the increase in quantity supplied of 5 per cent is smaller than the increase in price of 22 per cent, the supply curve is inelastic in this range.

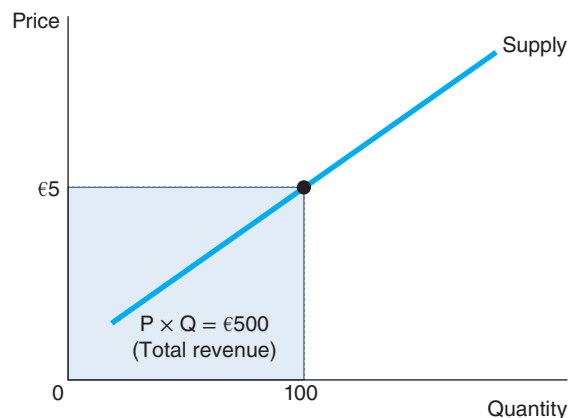
**Total Revenue and the Price Elasticity of Supply**

When studying changes in supply in a market we are often interested in the resulting changes in the **total revenue** received by producers. In any market, total revenue received by sellers is $P \times Q$, the price of the good times the quantity of the good sold. This is highlighted in Figure 4.8, which shows an upward sloping supply curve with an assumed price of €5 and a supply of 100 units. The height of the box under the demand curve is P and the width is Q . The area of this box, $P \times Q$, equals the total revenue received in this market. In Figure 4.8, where $P = €5$ and $Q = 100$, total expenditure is $€5 \times 100$ or €500.

total revenue the amount received by sellers of a good, computed as the price of the good times the quantity sold

FIGURE 4.8**The Supply Curve and Total Revenue**

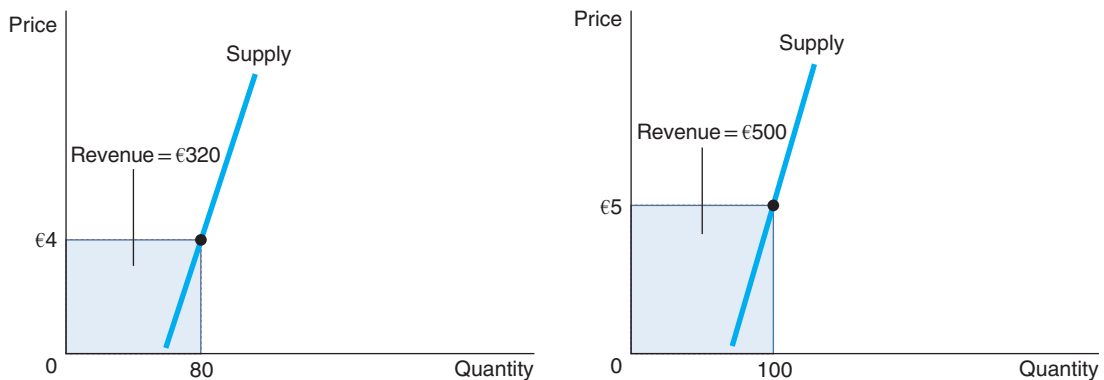
The total amount received by sellers equals the area of the box under the supply curve, $P \times Q$. Here, at a price of €5, the quantity supplied is 100 and the total revenue is €500.



How does total revenue change as one moves along the supply curve? The answer depends on the price elasticity of supply. If supply is inelastic, as in Figure 4.9, then an increase in the price which is proportionately larger causes an increase in total revenue. Here, an increase in price from €4 to €5 causes the quantity supplied to rise only from 80 to 100, and so total revenue rises from €320 to €500 (assuming the firm sold the additional supply).

FIGURE 4.9**How Total Revenue Changes When Price Changes: Inelastic Supply**

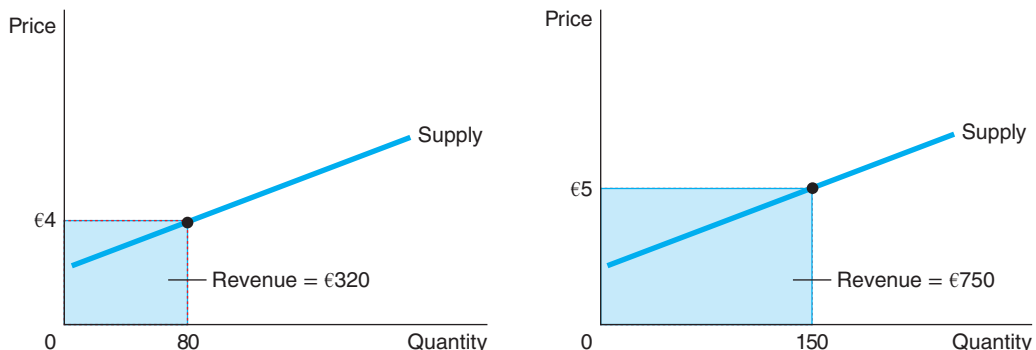
With an inelastic supply curve, an increase in the price leads to an increase in quantity supplied that is proportionately smaller. Therefore, total revenue (the product of price and quantity) increases. Here, an increase in the price from €4 to €5 causes the quantity supplied to rise from 80 to 100, and total revenue rises from €320 to €500.



If supply is elastic then a similar increase in price brings about a much larger than proportionate increase in supply. In Figure 4.10, we assume a price of €4 and a supply of 80 with total revenue of €320. Now a price increase from €4 to €5 leads to a much greater than proportionate increase in supply from 80 to 150 with total revenue rising to €750 – again, assuming the firm sold the additional supply.

FIGURE 4.10**How Total Revenue Changes When Price Changes: Elastic Supply**

With an elastic supply curve, an increase in the price leads to an increase in quantity supplied that is proportionately larger. Therefore, total revenue (the product of price and quantity) increases. Here, an increase in the price from €4 to €5 causes the quantity supplied to rise from 80 to 150, and total revenue rises from €320 to €750.



FYI



The Mathematics of Elasticity

We present this section for those who require some introduction to the maths behind elasticity. For those who do not need such a technical explanation, this section can be safely skipped without affecting your overall understanding of the concept of elasticity.

Point Elasticity of Demand

Figure 4.11 shows that the value for elasticity can vary at every point along a straight line demand curve. Point elasticity of demand allows us to be able to be more specific about the elasticity at different points. In the formula given below, the numerator (the top half of the fraction) describes the change in quantity in relation to the base quantity and the denominator the change in price in relation to the base price.

Price elasticity of demand

$$= \frac{\left(\frac{Q_2 - Q_1}{((Q_2 + Q_1)/2)} \right) \times 100}{\left(\frac{(P_2 - P_1)}{((P_2 + P_1)/2)} \right) \times 100}$$

If we cancel out the 100s in the above equation and re-write it a little more elegantly we get:

$$\text{Price elasticity of demand} = \frac{\Delta Q}{Q} \cdot \frac{P}{\Delta P}$$

Where the $\Delta Q = Q_2 - Q_1$

Re-arranging the above we get:

$$\text{Price elasticity of demand} = \frac{\Delta Q}{Q} \cdot \frac{P}{\Delta P}$$

There is no set order required to this equation so it can be re-written as:

$$\text{Price elasticity of demand} = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

Note that the expression $\Delta Q/\Delta P$ is the reciprocal of the slope of a linear demand curve. Look at the example in Figure 4.11 below:

Here we have two demand curves, D_1 and D_2 , given by the equations:

$$P = 20 - 5q(D_1)$$

and

$$Q = 5 - 0.25p(D_2)$$

For demand curve D_1 , the vertical intercept is 20 and the horizontal intercept is 4 and so the slope of the line D_1 is -5 .

For demand curve D_2 , the vertical intercept is 20 and the horizontal intercept is 5, the slope of the line D_2 is -4 .

To verify this let us take demand curve D_1 , if price were 10 then the quantity would be $10 = 20 - 5Q$.

Rearranging this gives:

$$20 - 10 = 5Q$$

Dividing both sides by 5 gives $Q = 2$

Looking at demand curve D_2 , if price were 10, then the quantity would be:

$$Q = 5 - 0.25(10)$$

$$Q = 5 - 2.5$$

$$Q = 2.5$$

Now let us assume that price falls from 10 to 5 in each case. The quantity demanded for D_1 would now be:

$$5 = 20 - 5Q$$

$$Q = 3$$

For D_2 the quantity would be:

$$Q = 5 - 0.25(5)$$

$$Q = 3.75$$

Representing this graphically for demand curve D_1 , we get the following as shown in Figure 4.12:

FIGURE 4.11

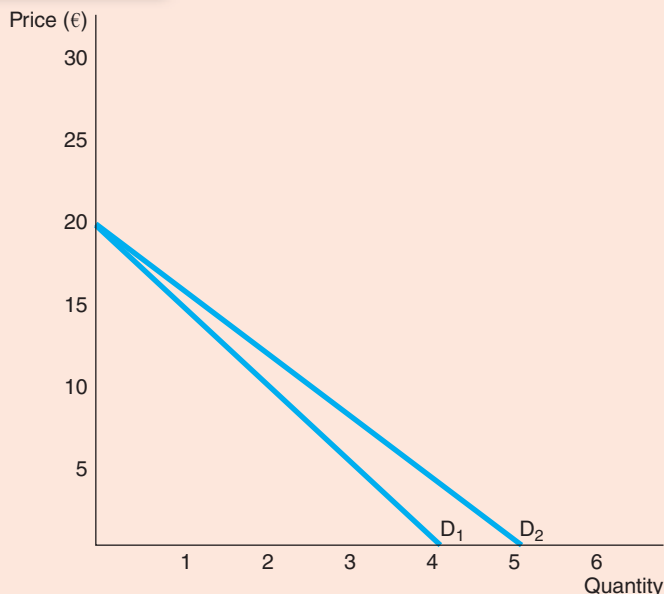


FIGURE 4.12



The slope of the line as drawn is:

$$\frac{\Delta P}{\Delta Q} = \frac{-5}{1} = -5$$

The slope is the same at all points along a linear demand curve. The price which we start with prior to a change will give different ratios at different points on the demand curve. Again, using demand curve D_2 , the ratio of P/Q at the initial price of 10 is $10/2 = 5$. At a price of 5, the ratio of P/Q given by the demand curve D_2 would be $5/3 = 1.67$.

Going back to our formula:

$$\text{Price elasticity of demand} = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

The first part of the equation ($\Delta Q/\Delta P$) is the reciprocal of the slope of the demand curve and the second part of the equation P/Q gives us a specific point on the demand curve relating to a particular price and quantity combination. Multiplying these two terms gives

us the price elasticity of demand at a particular point and so is referred to as *point elasticity of demand*.

The price elasticity of demand when price changes from 10 to 5 in demand curve D_2 above, would be:

$$\text{Price elasticity of demand} = \frac{1}{5} \cdot \frac{10}{2}$$

$$\text{Price elasticity of demand} = 1$$

If we were looking at a fall in price from 15 to 10 we would get:

$$\text{Price elasticity of demand} = \frac{1}{5} \cdot \frac{15}{1}$$

$$\text{Price elasticity of demand} = 3$$

And if looking at a price fall from 5 to 2.5 then we would get:

$$\text{Price elasticity of demand} = \frac{0.5}{5} \cdot \frac{10}{3}$$

$$\text{Price elasticity of demand} = 0.33$$

Calculus

The demand curve is often depicted as a linear curve but there is no reason why it should be linear and can be curvilinear. To measure

elasticity accurately in this case economists use calculus. The rules of calculus applied to a demand curve give a far more accurate measurement of the price elasticity of demand at a particular point.

For a linear demand function, the approximation to the point elasticity at the initial price and quantity is given by:

$$\frac{(Q_2 - Q_1)}{(P_2 - P_1)} \cdot \frac{P_1}{Q_1}$$

which gives exactly the same result as the point elasticity which is defined in terms of calculus and is given by:

$$\frac{dQ}{dP} \cdot \frac{P}{Q}$$

Point elasticity defined in terms of calculus gives a precise answer; all the other formulae are approximations of some sort. The formula looks similar but it must be remembered that what we are talking about in this instance is an infinitesimally small change in quantity following an infinitesimally small change in price. In the formula, dQ/dP is the derivative of a linear demand function. Given our basic linear equation of the form $Q = a - bP$, the power function rule gives dQ/dP as the coefficient of $P - (-b)$.

Take the following demand equation:

$$Q = 60 - 3P$$

To find the price elasticity of demand when price = 15 first of all we need to find Q :

$$Q = 60 - 3P$$

$$Q = 60 - 3(15)$$

$$Q = 15$$

We calculate dQ/dP as -3 .

Substitute this into the formula to get:

$$\text{Price elasticity of demand} = -3 \left(\frac{15}{15} \right)$$

$$\text{Price elasticity of demand} = -3$$

It is useful to remember that given an elasticity figure we can calculate the expected change in demand as a result of a change in price. For example, if the price elasticity of demand (ped) is given as 0.6 then an increase in price of 5 per cent will result in a fall in quantity demanded of 3 per cent. By using the inverse of the elasticity equation, for any given value of price elasticity of demand we can calculate how much of a price change is required to bring about a desired change in quantity demanded. Suppose that a government wanted to reduce the demand for motor vehicles as part of a policy to reduce congestion and pollution. What sort of price change might be required to bring about a 10 per cent fall in demand?

Assume that the price elasticity of demand for motor vehicles is 0.8. The inverse of the basic elasticity formula is:

$$\frac{1}{ped} = \frac{\% \Delta P}{\% \Delta Q}$$

Substituting our known values into the formula we get:

$$\frac{1}{0.8} = \frac{\% \Delta P}{10}$$

$$1.25 = \frac{\% \Delta P}{10}$$

$$\% \Delta P = 12.5$$

To bring about a reduction in demand of 10 per cent, the price of motor vehicles would have to rise by 12.5 per cent.

Other Elasticities

Income and cross elasticity of demand are all treated in exactly the same way as the analysis of price elasticity of demand above.

Point income elasticity would be:

$$\text{Income elasticity of demand} = \frac{\Delta Q}{\Delta Y} \cdot \frac{Y}{Q}$$

Using calculus:

$$\text{Income elasticity of demand} = \frac{dQ}{dY} \cdot \frac{Y}{Q}$$

For cross elasticity the formulas would be:

$$\text{Cross elasticity of demand} = \frac{\Delta Q_a}{\Delta P_b} \cdot \frac{P_b}{Q_a}$$

Where Q_a is the quantity demanded of one good, a , and P_b is the price of a related good, b (either a substitute or a complement). For a substitute $\frac{dQ_a}{dP_b} > 0$ and for a complement, $\frac{dQ_a}{dP_b} < 0$.

$$\text{Cross elasticity of demand} = \frac{dQ_a}{dP_b} \cdot \frac{P_b}{Q_a}$$

In Chapter 3 we saw that demand can be expressed as a multivariate function where demand is dependent on a range of variables which include price, incomes, tastes and so on. It is possible to calculate the elasticities of all these other factors using the same principles as those outlined above. In each case it is usual to calculate the elasticity with respect to a change in one of the variables whilst holding the others constant.

For example, take the demand equation $Q = 1,400 - 4P + 0.04Y$. This equation tells us that demand is dependent on the price and also the level of income.

From this equation we can calculate the price elasticity of demand and the income elasticity of demand. In this example we will use calculus to find both elasticities assuming $P = 50$ and $Y = 8,000$.

Given these values:

$$Q = 1,400 - 4(50) + 0.04(8,000)$$

$$Q = 1,400 - 200 + 320$$

$$Q = 1,520$$

With:

$$\frac{dQ}{dP} = -4$$

$$ped = -4 \left(\frac{50}{1,520} \right)$$

$$ped = -0.132$$

Given:

$$\frac{dQ}{dY} = 0.04$$

Income elasticity of demand

$$(\text{yed}) \frac{dQ}{dY} \cdot \frac{Y}{Q}$$

$$= 0.04 \left(\frac{8,000}{1,520} \right)$$

$$\text{yed} = 0.21$$

Now look at this demand equation:

$$Q_a = 100 - 8P_a - 6P_b + 4P_c + 0.015Y$$

This equation gives the relationship between demand and the prices of other goods labelled a , b and c respectively. We can use this to find the respective cross elasticities.

Assume that the price of good a is 20, the price of good b , 40, the price of good c , 80 and $Y = 20,000$.

Substituting these into our function gives:

$$Q_a = 100 - 8P_a - 6P_b + 4P_c + 0.015Y$$

$$Q_a = 100 - 8(20) - 6(40) + 4(80) + 0.015(20,000)$$

$$Q_a = 100 - 160 - 240 + 320 + 300$$

$$Q_a = 320$$

The change in demand of good a with respect to changes in the price of good b is given by:

$$\frac{dQ_a}{dP_b} = -6$$

The

$$\begin{aligned} \text{cross elasticity of demand} &= -6 \left(\frac{40}{320} \right) \\ &= -6(0.125) \\ &= -0.75 \end{aligned}$$

The relationship between goods a and b is that they are complements – a rise in the price of good b will lead to a fall in the quantity demanded of good a .

The change in demand of good a with respect to changes in the price of good c is given by:

$$\frac{dQ_a}{dP_c} = 4$$

$$\begin{aligned} \text{Cross elasticity of demand} &= 4 \left(\frac{80}{320} \right) \\ &= 4(0.25) \\ &= 1 \end{aligned}$$

In this case the relationship between the two goods is that they are substitutes – a rise in the price of good c would lead to a rise in the quantity demanded of good a .

Price Elasticity of Supply

Many of the principles outlined above apply also to the price elasticity of supply. The formula for the price elasticity of supply using the point method is:

$$\text{Price elasticity of supply} = \frac{\Delta Q_s}{\Delta P} \cdot \frac{P}{Q_s}$$

Using calculus:

$$\text{Price elasticity of supply} = \frac{dQ_s}{dP} \cdot \frac{P}{Q_s}$$

However, we need to note a particular issue with price elasticity of supply which relates to the graphical representation of supply curves.

This is summarized in the following:

- A straight line supply curve intersecting the y -axis at a positive value has a price elasticity of supply > 1
- A straight line supply curve passing through the origin has a price elasticity of supply $= 1$

- A straight line supply curve intersecting the x -axis at a positive value has a price elasticity of supply < 1

To see why any straight line supply curve passing through the origin has a price elasticity of supply of 1 we can use some basic knowledge of geometry and similar triangles.

Figure 4.13 shows a straight line supply curve S_1 passing through the origin. The slope of the supply curve is given by $\frac{\Delta P}{\Delta Q_s}$. We have high-

lighted a triangle, shaded green, with the ratio $\frac{\Delta P}{\Delta Q_s}$ relating to a change in

price of 7.5 and a change in quantity of 1. The larger triangle formed by taking a price of 22.5 and a quantity of 3 shows the ratio of the price and quantity at this point (P/Q). The two triangles formed by these are both classed as similar triangles – they have different lengths to their three sides but the internal angles are all the same. The ratio of the sides

must therefore be equal as shown by equation (1) below:

$$\frac{\Delta P}{\Delta Q_s} = \frac{P}{Q_s} \quad (1)$$

Given our definition of point elasticity of supply, if we substitute equation 1 into the formula and rearrange we get:

$$\text{Price elasticity of supply} = \frac{\Delta Q_s}{\Delta P} \frac{P}{Q_s}$$

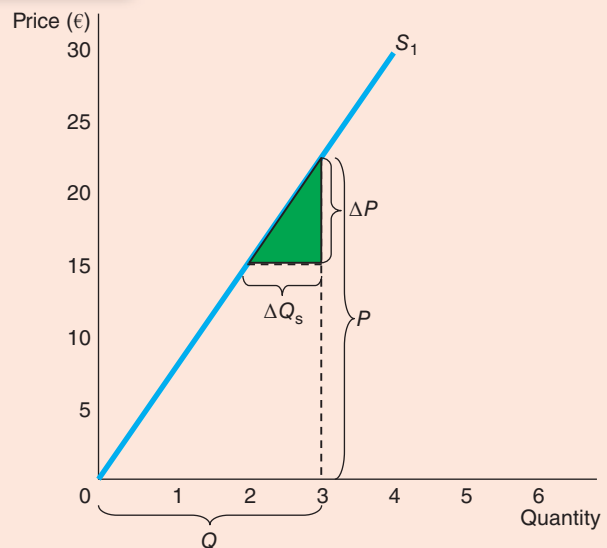
Therefore:

$$\text{Price elasticity of supply} = 1$$

Elasticity and Total Expenditure/Revenue

We have used the term ‘total expenditure’ in relation to the demand curve to accurately reflect the fact that demand is related to buyers and when buyers pay for products this represents expenditure. Many books use the term expenditure and revenue interchangeably and in this short section we are going to refer to revenue.

FIGURE 4.13



Total revenue is found by multiplying the quantity purchased by the average price paid. This is shown by the formula:

$$TR = P \times Q$$

Total revenue can change if either price or quantity, or both, change. This can be seen in Figure 4.14 where a rise in the price of a good from P_0 to P_1 has resulted in a fall in quantity demanded from Q_0 to Q_1 .

We can represent the change in price as ΔP so that the new price is $(P + \Delta P)$ and the change in quantity as ΔQ so that the new quantity is $(Q + \Delta Q)$ so TR can be represented thus:

$$TR = (P + \Delta P)(Q + \Delta Q)$$

If we multiply out this expression as shown then we get:

$$TR = (P + \Delta P)(Q + \Delta Q)$$

$$TR = PQ + P\Delta Q + \Delta PQ + \Delta P\Delta Q$$

In Figure 4.14, this can be seen graphically.

As a result of the change in price there is an additional amount of revenue shown by the blue rectangle ($Q\Delta P$). However, this is offset by the reduction in revenue caused by the fall in quantity demanded as a result of the change in price shown by the green rectangle ($P\Delta Q$). There is also an area indicated by the yellow rectangle which is equal to $\Delta P\Delta Q$. This leaves us with a formula for the change in TR as:

$$\Delta TR = Q\Delta P + P\Delta Q + \Delta P\Delta Q$$

Let us substitute some figures into our formula to see how this works in practice. Assume the original price of a product is 15 and the quantity demanded at this price is 750. When the price rises to 20 the quantity demanded falls to 500.

Using the equation:

$$TR = PQ + P\Delta Q + \Delta PQ + \Delta P\Delta Q$$

TR is now:

$$TR = 15(750) + 15(-250) + 5(750) + 5(-250)$$

$$TR = 10,000$$

The change in TR is:

$$\Delta TR = Q\Delta P + P\Delta Q + \Delta P\Delta Q$$

$$\Delta TR = 750(5) + 15(-250) + 5(-250)$$

$$\Delta TR = 3,750 - 3,750 - 1,250$$

$$\Delta TR = -1,250$$

In this example the effect of the change in price has been negative on TR . We know from our analysis of price elasticity of demand that this means the percentage change in quantity demanded was greater than the percentage change in price – in other words, price elasticity of demand must be elastic at this point (>1). For the change in TR to be positive, therefore, the price elasticity of demand must be <1 .

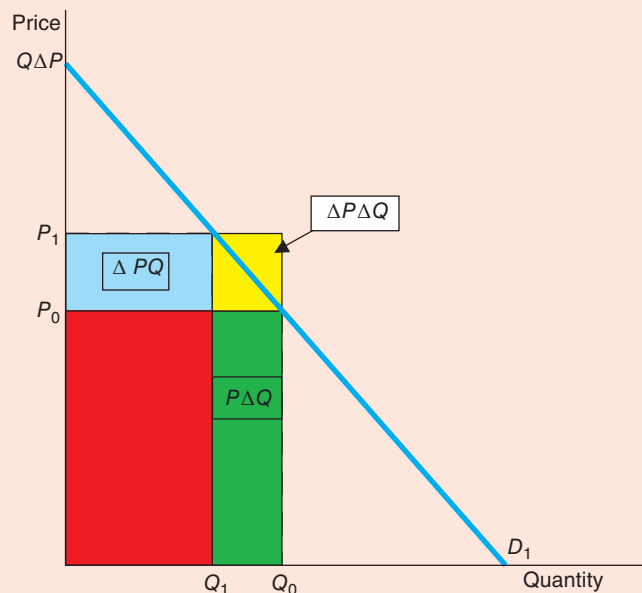
We can express the relationship between the change in TR and price elasticity of demand as an inequality as follows:

Price elasticity of demand

$$= \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q} > 1$$

FIGURE 4.14

The original TR is found by multiplying the original price (P_0) by the original quantity (Q_0) and is shown by the red + green rectangles.



When price increases, revenue decreases if price elasticity of demand meets this inequality.

Equally, for a price increase to result in a rise in revenue *ped* must meet the inequality below:

Price elasticity of demand

$$= \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q} < 1$$

APPLICATIONS OF SUPPLY AND DEMAND ELASTICITY

Why is it the case that travel on the trains at some times during the day is a different price than at other times? Why, despite the increase in productivity in agriculture, have farmer's incomes gone down on average over recent years? At first, these questions might seem to have little in common. Yet both questions are about markets and all markets are subject to the forces of supply and demand. An understanding of elasticity is a key part of the answer to these and many other questions.

Why Does the Price of Train Travel Vary at Different Times of the Day?

In many countries the price of a train journey varies at different times during the day and the week. A ticket for a seat on a train from Birmingham to London between 6.00am and 9.00am is around £80 (€99), whereas for the same journey leaving at midday the price is between £10 and £20 (€12 and €24). Train operators know that the demand for rail travel between 6.00am and 9.00pm is higher than during the day time but they also know that few commuters have choices about when they have to arrive at work or to meetings, conferences and so on. An individual can use other forms of transport such as their car or a coach but the train is often very convenient, so the amount of substitutes is considered low. The price elasticity of demand for train travel early in the morning, therefore, is relatively low compared to at midday. In the morning, train operators know that seats on trains will be mostly taken and there will be very few left empty, whereas during the day it is much more likely that trains will be running with empty seats. Knowing that there is a different price elasticity of demand means that train operators can maximize revenue at these different times by charging different prices.

Figure 4.15 shows the situation in the market for train travel. Panel (a) shows the demand and supply for tickets between Birmingham and London between 6.00am and 9.00am. The demand curve D_1 is very steep indicating that the price elasticity of demand is relatively low. At a price of £80, 1,000 tickets are bought and as a result the total revenue for the train operator is £80,000.

Panel (b) shows a demand curve D_2 with a similar supply curve. *Ceteris paribus*, the train operator has the same number of trains available at all times during the day but notice that the supply curve is relatively steep and therefore inelastic because although the operator has some flexibility to increase the number of trains available and thus seats for passengers, there is a limit as to how far the capacity can be varied throughout the day.

If the train operator charged a price of £80 after 9.00am, the demand for tickets would be very low at 100. Total revenue, therefore would be £8,000 and there would be many seats left empty. This is because the train operator effectively faces a different market during the day. Those who travel by train at this time may have a choice – they might be travelling for leisure or to see friends and they do not *need* to travel by train unlike those in the morning who have to get to work at a certain time. These passengers are price sensitive – charge too high a price and they will not choose to travel by train but offer a price that these passengers see as being attractive and which to them represents value for money and they may choose to buy a train ticket.

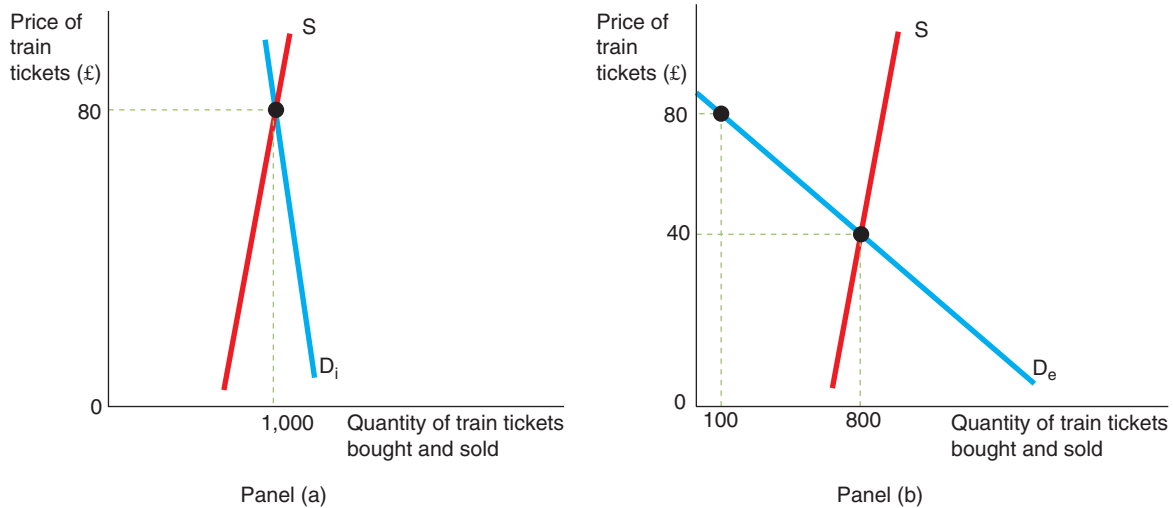
If the train operator, therefore, set the price for train tickets after 9.00am at £40 the demand for train tickets would be 800 and total revenue would be £32,000. If we assume that train operators are acting rationally then they would prefer to generate revenue of £32,000 rather than £8,000 and so it would be more sensible for them to charge a lower price to capture these more price sensitive passengers.

SELF TEST How might a drought that destroys half of all farm crops be good for farmers? If such a drought is good for farmers, why don't farmers destroy their own crops in the absence of a drought?

FIGURE 4.15**Price Sensitivity in the Passenger Train Market**

Panel (a) represents the market for train travel between 6.00am and 9.00am between two major cities. The demand for train travel at this time is relatively price inelastic – passengers are insensitive to price at this time because they have few alternatives and have to get into work and to meetings. The train operators generate revenue of £80,000 by selling 1,000 tickets at £80 each.

Panel (b) shows the market after 9.00am. Train operators face a different demand curve at this time and passengers are more price sensitive. If the train operator continued to charge £80, demand would be just 100 and revenue would be £8,000. If the train operator reduces the price to £40, demand would be 800 and so the total revenue would be £32,000.

**Why Have Farmer's Incomes Fallen Despite Increases in Productivity?**

In many developed countries, agricultural production has increased over the last 50 years. One of the reasons is that farmers are able to use more machinery, and advances in science and technology have meant that productivity – the amount of output per acre of land, has increased. Assume a farmer has 1,000 acres of land and grows wheat. Twenty years ago each acre of land yielded an average of 2.0 tonnes of wheat. We say an average because output can be dependent on factors outside the farmer's control such as the weather, pests, diseases and so on. Assume that the price of wheat is €200 per tonne. Twenty years ago the average income for our farmer would be 2,000 tonnes \times €200 = €400,000.

Ceteris paribus, if productivity increases meant that average output per acre was now 3.0 tonnes per acre, income would rise to €600,000.

However, this assumes other things are equal. Research suggests that the demand for food is relatively price inelastic and also income inelastic. Over the 20-year period we are assuming in our analysis, the demand for wheat may only have risen by a relatively small amount and also is price and income inelastic. People may earn more money now than they did 20 years ago but Engel's law (see Chapter 2) tells us that the proportion of their income devoted to food tends to decline.

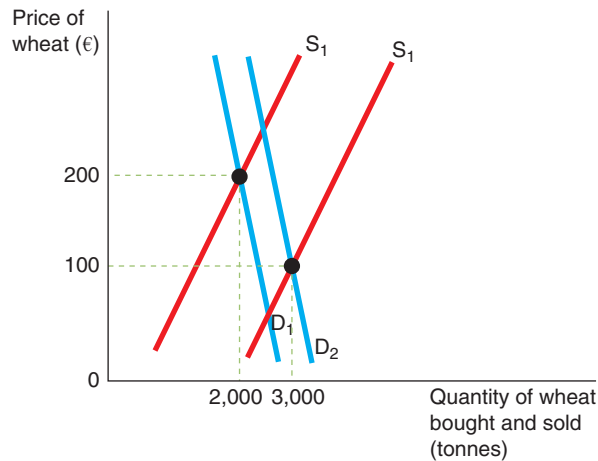
Figure 4.16 shows a representation of this situation. In the first time period the supply curve representing output per acre of 2.0 tonnes intersects the demand curve D_1 at a price of €200 per tonne giving the farmer an income of €400,000.

20 years later, the productivity improvements at the farm see the supply curve shifting to S_2 representing an average output per acre of 3.0 tonnes. However, over the 20-year period, demand has increased but only by a small amount as people spend a smaller proportion of their income on food as they get richer. The fact that food is relatively price inelastic is indicated by the relatively steep demand curve and the result is that the market price has fallen to €100 per tonne with the farmer now selling 3,000 tonnes. The farmer's income has fallen to €300,000.

FIGURE 4.16**The Effect of Increases in Demand and Supply of Wheat on Farm Incomes**

Twenty years ago, the supply of wheat is represented as supply curve S_1 with output per acre at 2.0 tonnes per acre. The demand for wheat at that time is represented as demand curve D_1 . If the market price of wheat is €200 per tonne, the farmer's income is €400,000.

The output of wheat per acre rises with increases in productivity and as a result the supply of wheat today increases and is represented by supply curve S_2 with output per acre now 3.0 tonnes per acre. However, as demand is both price and income inelastic, the demand for food has increased only slightly in that 20-year period; the new demand curve is shown as D_2 . The combination of a considerable rise in supply and only a small rise in demand means farmers get a lower price per tonne and income is actually lower.

**IN THE NEWS****The Future Price of Oil and Elasticity**

At the time of writing the price of oil on world markets is hovering around \$100 a barrel. The future price of oil will be dependent on a number of factors, but the concept of elasticity will have some relevance to all the questions which follow this article.

IMF Study: Peak Oil could do Serious Damage to the Global Economy

The world isn't going to run out of oil anytime soon. But there's still concern among various geologists and analysts that our oil supply won't grow as quickly or as easily as it used to. We'll have to resort to harder-to-drill oil to satisfy our crude habits. That would push prices up.

And high oil prices could act as a drag on growth.

This, at any rate, is the basic idea behind 'peak oil'. And there's some reason for worry. Between 1981 and 2005, world oil production grew at a steady pace of about 1.8 per cent per year. All was well. But starting around 2005, oil production appeared to plateau. And, since demand for oil kept rising, especially in countries like China and India, that

caused prices to soar. Oil doesn't get much cheaper than \$100 per barrel these days. And that, some economists worry, has acted as a drag on growth around the world.

So how bad would it be if peak oil was really upon us? That's a question that two International Monetary Fund (IMF) economists try to tackle in a new working paper, 'Oil and the World Economy: Some Possible Futures'. The authors, Michael Kumhof and

Dirk Muir, don't make any definitive predictions about how the oil supply will evolve. Rather, they try to model a number of different scenarios in which oil *does* become scarcer and the world tries to adapt.

The paper itself offers an interesting look at how the world might cope with higher oil prices, so let's take a look at the various scenarios:

1) Oil production grows very slowly or plateaus. This is the baseline scenario that Kumhof and Muir use. They assume that oil grows by about 1 percentage point less each year than its historical average. So, let's say, oil grows at a steady 0.8 per cent per year rather than the 1.8 per cent annual average between 1981 and 2005. This isn't a temporary oil disruption like the one we saw in 2008. It's a persistent, long-term supply shock.

What would happen? Oil prices, the IMF model suggests, would gradually double in 10 years and quadruple over 20 years. Regions that import oil on net, such as Europe and the United States, would see a small hit to growth – about

As supplies of oil start to diminish, oil producers will find it harder and more expensive to find new supplies of oil. Traditional methods of extraction using these so-called 'nodding horse' pumping units may become a thing of the past.



0.2 to 0.4 percentage points each year. Countries that export, like Saudi Arabia, would get a lot wealthier.

2) Oil production grows at a slower rate, but the world adapts fairly easily. In this scenario, oil production declines, but countries start switching to electric cars or fuelling their vehicles with natural gas. Vehicles and manufacturers become more efficient. In economist terms, the 'elasticity' of demand quickly increases.

Under this scenario, the United States and Europe take just a small hit to growth, about 0.1 to 0.2 percentage points per year. Japan and Asia actually get a boost to their economy, since they can adapt to higher oil prices *and* export more stuff to oil-producing countries in the Middle East. All told, this is a fairly happy outcome.

3) Oil production grows at a slower rate, but the world can't find substitutes. As the IMF authors note, it's not assured that the world can quickly adapt to steadily increasing oil prices. Oil is, after all,

quite valuable and hard to replace. Electric cars may not catch on. It's tough to build infrastructure for natural-gas vehicles. The chemical industry might struggle to find substitutes for oil as feedstock. The oil substitutes that result turn out to be lower quality. In this scenario, wealthy regions like Europe, the United States and Japan take an annual gross domestic product (GDP) hit of 0.4 per cent to 0.6 per cent. That starts to hurt.

4) Oil turns out to be far more important than most economists had assumed. The Energy Information Administration estimates that petroleum purchases make up just 3.5 per cent of the US economy. Looked at from that angle, expensive oil shouldn't do too much damage. But, the IMF authors note, several books and articles have pointed out that this understates how crucial oil is to the functioning of a modern economy. Many key technologies contain materials or use fuels derived from crude.

If, in fact, oil is much more important than many economic modellers have assumed, then the blow to growth from even a modest plateau in oil could be quite large – lowering growth rates by up to 1.2 percentage points over the next two decades.

5) Oil production starts shrinking rapidly. This is the Domesday scenario. Some studies have suggested that global oil production is currently on a plateau and will soon start shrinking by around 2 per cent per year. Existing wells will dry up. The world will increasingly rely on oil from places that are more expensive to develop, such as Canada's tar sands. What happens then?

Nothing good. According to the IMF's modelling, prices could increase by 800 per cent over two decades. Growth rates in Europe and the United

States would be reduced by at least a full percentage point – and much more if oil turns out to be more important than we thought. ‘Relative price changes of this magnitude would be unprecedented’, the authors note, ‘and would almost certainly have non-linear effects on GDP that the model is not able to capture adequately’. Yikes.

In any case, these scenarios aren’t easy to model – especially since nations might respond in unpredictable ways. (If crude output started shrivelling, some oil producers could start restricting exports.

Or fuel subsidies could affect demand elasticity.) All told, however, the IMF authors say it’s quite possible that a decent-sized decline in oil production could have ‘dramatic’ effects that could prove very, very difficult for the world to adjust to.

Questions

- 1 What do you think is going to happen to the price elasticity of supply for oil in the next 50 years? Explain.
- 2 Assuming the supply of oil did plateau in 2005, use the concept

of price elasticity to explain why increasing demand for oil might lead to prices rising significantly in the years to come.

- 3 If oil prices did rise in line with Scenario 1 in the article, why would countries like Saudi Arabia become much wealthier?
- 4 In Scenario 2, why is the outcome of the decline in oil production dependent on the ease with which the world adapts?
- 5 What effect might fuel subsidies have on the price elasticity of demand for oil?

SUMMARY

- The price elasticity of demand measures how much the quantity demanded responds to changes in the price. Demand tends to be more elastic if close substitutes are available, if the good is a luxury rather than a necessity, if the market is narrowly defined or if buyers have substantial time to react to a price change.
- The price elasticity of demand is calculated as the percentage change in quantity demanded divided by the percentage change in price. If the elasticity is less than 1, so that quantity demanded moves proportionately less than the price, demand is said to be inelastic. If the elasticity is greater than 1, so that quantity demanded moves proportionately more than the price, demand is said to be elastic.
- The price elasticity of supply measures how much the quantity supplied responds to changes in the price. This elasticity often depends on the time horizon under consideration. In most markets, supply is more elastic in the long run than in the short run.
- The price elasticity of supply is calculated as the percentage change in quantity supplied divided by the percentage change in price. If the elasticity is less than 1, so that quantity supplied moves proportionately less than the price, supply is said to be inelastic. If the elasticity is greater than 1, so that quantity supplied moves proportionately more than the price, supply is said to be elastic.
- Total revenue, the total amount received by sellers for a good, equals the price of the good times the quantity sold. For inelastic demand curves, total revenue rises as price rises. For elastic demand curves, total revenue falls as price rises.
- The income elasticity of demand measures how much the quantity demanded responds to changes in consumers’ income. The cross-price elasticity of demand measures how much the quantity demanded of one good responds to changes in the price of another good.

QUESTIONS FOR REVIEW

- 1 Define the price elasticity of demand and the income elasticity of demand.
- 2 List and explain some of the determinants of the price elasticity of demand. Think of some examples to use to illustrate the factors you cover.
- 3 If the elasticity is greater than 1, is demand elastic or inelastic? If the elasticity equals 0, is demand perfectly elastic or perfectly inelastic?
- 4 How is the price elasticity of supply calculated? Explain what this measures.

- 5 Is the price elasticity of supply usually larger in the short run or in the long run? Why?
- 6 What are the main factors that affect the price elasticity of supply? Think of some examples to use to illustrate the factors you cover.
- 7 A business person reads that the price elasticity of demand for the product they sell is 1.25. If the person wishes to increase revenue should they increase or reduce price? Explain.
- 8 In the very short run, the price elasticity of demand for tobacco products is very low but over the medium term is less so. Explain.
- 9 What do we call a good whose income elasticity is less than 1?
- 10 What factors might affect the price elasticity of supply for a commodity such as rubber in the short run and the long run? Explain.

PROBLEMS AND APPLICATIONS

- 1 Seafront properties along the south coast of France have an inelastic supply, and cars have an elastic supply. Suppose that a rise in population doubles the demand for both products (that is, the quantity demanded at each price is twice what it was).

- a. What happens to the equilibrium price and quantity in each market?
- b. Which product experiences a larger change in price?
- c. Which product experiences a larger change in quantity?
- d. What happens to total consumer spending on each product?

- 2 Because better weather makes farmland more productive, farmland in regions with good weather conditions is more expensive than farmland in regions with bad weather conditions. Over time, however, as advances in technology have made all farmland more productive, the price of farmland (adjusted for overall inflation) has fallen. Use the concept of elasticity to explain why productivity and farmland prices are positively related across space but negatively related over time.

- 3 For each of the following pairs of goods, which good would you expect to have more elastic demand and why?

- a. Required textbooks or mystery novels.
- b. Downloads of tracks by Jay-Z or downloads of hip-hop music in general.
- c. Heating oil during the next six months or heating oil during the next five years.
- d. Lemonade or water.

- 4 Suppose that business travellers and holidaymakers have the following demand for airline tickets from Birmingham to Naples:

| Price (€) | Quantity demanded (business travellers) | Quantity demanded (holidaymakers) |
|-----------|--|--------------------------------------|
| 150 | 2,100 | 1,000 |
| 200 | 2,000 | 800 |
| 250 | 1,900 | 600 |
| 300 | 1,800 | 400 |

- a. As the price of tickets rises from €200 to €250, what is the price elasticity of demand for (i) business travellers and (ii) holidaymakers? (Use the midpoint method in your calculations.)
 - b. Why might holidaymakers have a different elasticity from business travellers?
- 5 Suppose that your demand schedule (per year) for DVDs is as follows:

| Price (€) | Quantity demanded (income = €10,000) | Quantity demanded (income = €12,000) |
|-----------|---|---|
| 8 | 40 | 50 |
| 10 | 32 | 45 |
| 12 | 24 | 30 |
| 14 | 16 | 20 |
| 16 | 8 | 12 |

- a. Use the midpoint method to calculate your price elasticity of demand as the price of DVDs increases from €8 to €10 if (i) your income is €10,000, and (ii) your income is €12,000.
- b. Calculate your income elasticity of demand as your income increases from €10,000 to €12,000 if (i) the price is €12, and (ii) the price is €16.

- 6 Karen receives an energy bill and on looking through it notes that she is charged a different rate for gas and electricity during the day compared to the night time. Explain to Karen the likely reason for this difference in tariff.
- 7 Consider public policy aimed at smoking.
- Studies indicate that the price elasticity of demand for cigarettes is about 0.4. If a packet of cigarettes is currently priced at €6 and the government wants to reduce smoking by 20 per cent, by how much should it increase the price?
 - If the government permanently increases the price of cigarettes, will the policy have a larger effect on smoking one year from now or five years from now?
 - Studies also find that teenagers have a higher price elasticity of demand for cigarettes than do adults. Why might this be true?
- 8 Many governments impose a duty (tax) on petrol. Why do you think governments choose to tax petrol?
- 9 Suppose that there is severe flooding in a region in which there is a high concentration of wheat farmers.
- Farmers whose crops were destroyed by the floods were much worse off, but farmers whose crops were not destroyed benefited from the floods. Why?
 - What information would you need about the market for wheat to assess whether farmers as a group were hurt or helped by the floods?
- 10 Explain why the following might be true: a drought around the world raises the total revenue that farmers receive from the sale of grain, but a drought only in France reduces the total revenue that French farmers receive.

5

BACKGROUND TO DEMAND: THE THEORY OF CONSUMER CHOICE

Chapters 3 and 4 have looked at the basic theory of markets and how magnitudes of change in demand and supply can be measured using the concept of elasticity. In the next two chapters we will look in more detail at how the demand and supply curves are derived. In this chapter we look in more detail at the behaviour of consumers.

In Chapter 3 we stated the law of demand which is that the amount consumers wish to purchase is negatively related to price. On the face of it this seems intuitive. Behind this law, however, is a model of how consumers behave which is dependent on a number of assumptions. We will begin by looking at this model, then look at some limitations of the model and how recent research is shedding new light on consumer behaviour and challenging this model.

THE STANDARD ECONOMIC MODEL

When you walk into a shop, you are confronted with thousands of goods that you might buy. Of course, because your financial resources are limited, you cannot buy everything that you want. The assumption is, therefore, that you consider the prices of the various goods being offered for sale and buy a bundle of goods that, given your resources, best suits your needs and desires. In economic terminology, you are seeking to maximize your utility subject to the constraint of a limited income.

This model is called the classical theory of consumer behaviour or the *standard economic model (SEM)* and is fundamentally based on an assumption that humans behave rationally when making consumption choices.

The standard economic model provides a theory of consumer choice which provides a more complete understanding of demand. One of the *Ten Principles of Economics* discussed in Chapter 1 is that people face trade-offs. The theory of consumer choice examines the trade-offs that people face in their role as consumers. When a consumer buys more of one good, he can afford less of other goods. When he spends more time enjoying leisure and less time working, he has lower income and can afford less consumption. When he spends more of his income in the present and saves less of it, he must accept a lower level of consumption in the future. The theory of consumer choice examines how consumers facing these trade-offs make decisions and how they respond to changes in their environment.

When making these trade-offs, there are a number of assumptions that are made about consumers. These include:

- Buyers (or economic agents as they are sometimes referred to) are rational
- More is preferred to less
- Buyers seek to maximize their utility
- Consumers act in self-interest and do not consider the utility of others.

Value

A key concept in consumer behaviour and across many other areas of economics is value. Value can be regarded as a threshold concept – an understanding of the concept of value helps to provide a transformed

understanding of economics. Value is a subjective term – what one individual thinks represents value is often different to that of another individual. Value can be seen as the worth to an individual of owning an item represented by the satisfaction derived from its consumption. In broad terms, consumption in this case does not just mean the final consumer. Value can be related to the purchase of a product which is a gift or something used by a business for production.

Classical economists used the term **utility** to refer to the satisfaction derived from consumption.

utility the satisfaction derived from the consumption of a certain quantity of a product

Utility is an ordinal concept, what this means is that it can be used as a way of ranking but cannot have any meaningful arithmetic operations performed on it. A measure of utility, therefore, can be used to represent consumer choices in some order but that order tells us nothing about the differences in the values we use. For example, if a group of five people were asked to rank different movies in order of preference using a 10-point scale (with each point referred to as a util) we might be able to conclude that movie 5 was the most popular, followed by movie 3 and movie 8. If, however, person 1 ranked movie 5 at 10 utils, whilst person 2 ranked the same movie as 5 utils, we cannot say that person 1 values movie 5 twice as much as person 2, only that they place it higher in their preferences. Value can be measured, therefore, by ranking but there are limitations to such ranking.

One way in which we can overcome this limitation is to look at value in terms of the amount consumers are prepared to pay to secure the benefits of consuming the product. This is called the *willingness to pay principle*. It is highly likely that at some point in your life you will have said something like ‘I wouldn’t have that if you paid me’ or similar. How much of our limited income we are prepared to pay is a reflection of the value we put on acquiring a good. It might not tell us much about the satisfaction from actually consuming the good (the buyer, as we have seen, might not be the final consumer) but it does give some idea of value.

For example, two friends, Alexa and Monique are in a store looking at a pair of shoes. Alexa picks up a pair of leopard print high heel shoes priced at €75. Monique looks at her friend and frowns – why on earth is she thinking of buying those? No way would Monique pay that sort of money for such an awful pair of shoes. A discussion ensues about the shoes; clearly there is a difference of opinion on them and thus how much it is ‘worth’ it to own this particular pair of shoes. If Alexa buys the shoes, then the value to her

Would you pay €75 for these shoes? The answer might depend on your gender for a start, then it might depend on your tastes. Some people might think at €75 these shoes represent a bargain; others a waste of money. What do you think and why?



must be at least €75 because that is what she has to give up in money terms in order to acquire them. It may be that Alexa would have been prepared to pay much more for the shoes, in which case she is getting some additional benefit which she is not paying for. Economists call this consumer surplus and we will look in more detail at this later in the book. Alexa sticks to her guns and buys the shoes; Monique leaves the store baffled at her friend’s purchasing decision. Monique clearly feels that giving up €75 to buy those shoes was a ‘waste of money’.

The amount buyers are prepared to pay for a good, therefore, tells us something about the value they place on it. It is not just the amount of money we hand over that reflects value but what that amount of money could have bought. This highlights one of the *Ten Principles of Economics* – the cost of something is what you have to give up to get it, the opportunity cost. We could make a reasonable assumption that Monique believed there was a way in which she could allocate €75 to get more value – in other words, the alternative that €75 could buy (whatever that might be) represented greater value than purchasing those shoes.

SELF TEST Think about your purchasing decisions. Sometimes you will think you have got a ‘bargain’ and other times you turn away from a purchase because you think it will be a ‘waste of money’. In economics, what do you think is the difference and why?

THE BUDGET CONSTRAINT: WHAT THE CONSUMER CAN AFFORD

One of the assumptions the SEM makes is that more is preferred to less. Most people would like to increase the quantity or quality of the goods they consume – to take longer holidays, drive fancier cars or buy a bigger house. People consume less than they desire because their spending is *constrained*, or limited, by their income.

To keep things simple, let us use a model which examines the decisions facing a consumer who buys only two goods: cola and pizza. Of course, real people buy thousands of different kinds of goods. Yet using this model greatly simplifies the problem without altering the basic insights about consumer choice.

Suppose that the consumer has an income of €1,000 per month and that he spends his entire income each month on cola and pizza. The price of a litre of cola is €2 and the price of a pizza is €10.

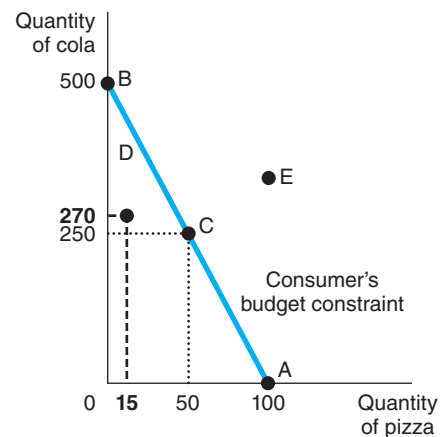
The table in Figure 5.1 shows some of the many combinations of cola and pizza that the consumer can buy. The first row in the table shows that if the consumer spends all his income on pizza, he can eat 100 pizzas during the month, but he would not be able to buy any cola at all. The second line shows another possible consumption bundle: 90 pizzas and 50 litres of cola. And so on. Each consumption bundle in the table uses up the consumer's income – exactly €1,000.

FIGURE 5.1

The Consumer's Budget Constraint

The budget constraint shows the various bundles of goods that the consumer can afford for a given income. Here the consumer buys bundles of cola and pizza. The table and graph show what the consumer can afford if his income is €1,000, the price of cola is €2 and the price of pizza is €10.

| Litres of cola | Spending on cola (€) | Number of pizzas | Spending on pizza (€) | Total spending (€) |
|----------------|----------------------|------------------|-----------------------|--------------------|
| 0 | 0 | 100 | 1,000 | 1,000 |
| 50 | 100 | 90 | 900 | 1,000 |
| 100 | 200 | 80 | 800 | 1,000 |
| 150 | 300 | 70 | 700 | 1,000 |
| 200 | 400 | 60 | 600 | 1,000 |
| 250 | 500 | 50 | 500 | 1,000 |
| 300 | 600 | 40 | 400 | 1,000 |
| 350 | 700 | 30 | 300 | 1,000 |
| 400 | 800 | 20 | 200 | 1,000 |
| 450 | 900 | 10 | 100 | 1,000 |
| 500 | 1,000 | 0 | 0 | 1,000 |



The graph of this data is given in Figure 5.1. The line connecting points A to B is called the **budget constraint** and shows the consumption bundles that the consumer can afford given a specified income. Five points are marked on this figure. At point A, the consumer buys no cola and consumes 100 pizzas. At point B, the consumer buys no pizza and consumes 500 litres of cola. At point C, the consumer buys 50 pizzas and 250 litres of cola. At point C the consumer spends an equal amount (€500) on cola and pizza. Point D is inside the budget constraint.

budget constraint the limit on the consumption bundles that a consumer can afford

The consumer can afford to buy any combination inside the budget constraint, in this example, point D shows a combination of 270 litres of cola and 15 pizzas. Point E is outside the budget constraint. No points outside the budget constraint are possible – the consumer does not have the income to be able to afford any combination of pizza and cola to the right of the budget constraint. Of course, these are only four of

the many combinations of cola and pizza that the consumer can choose given a specified income. All the points on and inside the line from A to B are possible. In this case, it shows the trade-off between cola and pizza that the consumer faces.

For example, assume the consumer is at point A, consuming 100 pizzas and zero cola. If the consumer wants to purchase a drink to go with his pizza he must give up some pizza in order to buy some cola – he has to trade-off the benefits of consuming cola against the benefits foregone of reducing consumption of pizza. We can quantify this trade-off. If the consumer moves to point C then he has to forego the benefits that 50 pizzas would provide to gain the benefits that 250 litres of cola would bring. The consumer would have to make a decision about whether it is worth giving up those 50 pizzas to get the benefits of the cola.

The slope of the budget constraint measures the rate at which the consumer can trade one good for the other. Remember, the slope between two points is calculated as the change in the vertical distance divided by the change in the horizontal distance ('rise over run'). From point A to point B, the vertical distance is 500 litres, and the horizontal distance is 100 pizzas. Because the budget constraint slopes downwards, the slope is a negative number – this reflects the fact that to get one extra pizza, the consumer has to *reduce* his consumption of cola by five litres. In fact, the slope of the budget constraint (ignoring the minus sign) equals the *relative price* of the two goods – the price of one good compared to the price of the other. A pizza costs 5 times as much as a litre of cola, so the opportunity cost of a pizza is 5 litres of cola. The budget constraint's slope of 5 reflects the trade-off the market is offering the consumer: 1 pizza for 5 litres of cola.

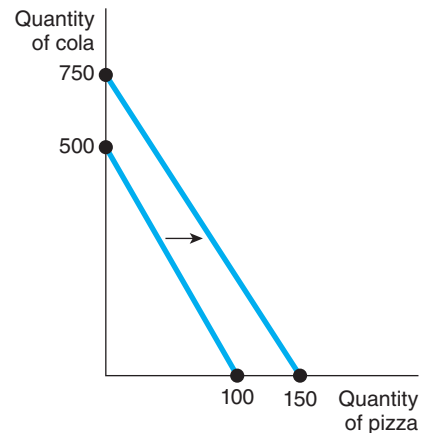
The opportunity cost of buying one extra pizza is five litres of cola. It is useful to use a rule of thumb (a heuristic) here; the opportunity cost of a good on the horizontal axis (pizza in our example) is the slope of the budget constraint (5 in this example). What is the opportunity cost of 1 extra litre of cola in our example? The opportunity cost of the good on the vertical axis is the inverse of the slope of the budget constraint which in this case is $\frac{1}{5}$ or 0.2. To acquire 1 extra litre of cola the consumer has to sacrifice one-fifth of a pizza. In some examples of budget constraints you may encounter, you might find the opportunity cost calculation does not make much sense. If the two goods being considered were cola and tins of soup you could not ask the shop to chop up the tin of soup into five! What is important to remember is that the slope is related to the ratio of prices of the goods being considered (the relative prices), $\frac{P_x}{P_y}$ where P_y is the price of the good on the vertical axis and P_x is the price of the good on the horizontal axis. In our example the ratio of the prices is $\frac{10}{2}$ which when simplified becomes 5.

SELF TEST Draw a budget constraint for a person with income of €5,000 if the price of food is €10 per unit and the price of leisure is €15 (per hour). What is the slope of this budget constraint? What is the opportunity cost of an extra hour of leisure in terms of food?

A Change in Income

Over time people's incomes change – sometimes they earn more but sometimes they earn less such as if they are made redundant, for example. If our consumer gets a pay rise and now earns €1,500 per month, he can now afford to buy more of both pizza and cola assuming the prices of these two goods do not change. The effect on the budget constraint is to cause it to shift to the right, as shown in Figure 5.2. If the consumer devoted all of his income to buying cola he could now buy 750 litres of cola compared to 500 litres when his income was €1,000 per month. If the consumer devoted all his income to buying pizza he could not afford to buy 150 pizzas a month compared to 100 when his income was €1,000 per month. Any point along the new budget constraint shows that the consumer can now buy more of both goods. If the consumer's income was to fall to €500 per month because he lost his job, for example, then the budget constraint would shift to the left, indicating that the consumer could now afford to buy less of both goods with his income.

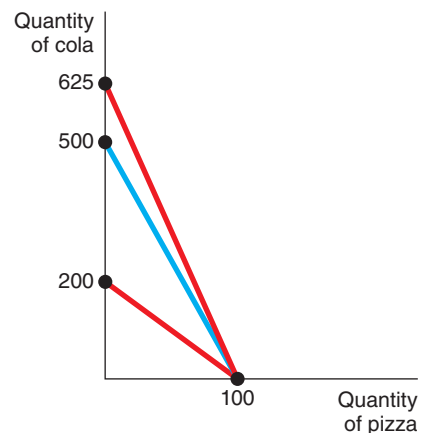
Notice, however, that whilst the budget constraint in Figure 5.2 has shifted to the right, the slope is still the same. This is because the prices of the two goods have not changed. What happens to the budget constraint if one or more of the prices of cola and pizza changes? This is what we look at next.

FIGURE 5.2**The Effect on the Budget Constraint of a Change in Income****A Change in Prices**

A Change in the Price of Cola Assume the consumer's income is €1,000 per month and the price of cola is €2 per litre and the price of pizza is €10. The budget constraint would look like that shown in Figure 5.1. Now assume that the price of cola rises to €5 per litre. The consumer would now only be able to afford to buy 200 litres of cola if he devoted all his income to cola. The budget constraint would, therefore, pivot inwards as shown in Figure 5.3.

FIGURE 5.3**A Change in the Price of Cola**

If the price of cola rises from €2 per litre to €5 per litre, the consumer can now afford to buy less cola with his income. The budget line pivots inwards and if the consumer devoted all his income to cola he would only be able to buy 200 litres compared to 500 litres before the price changed.



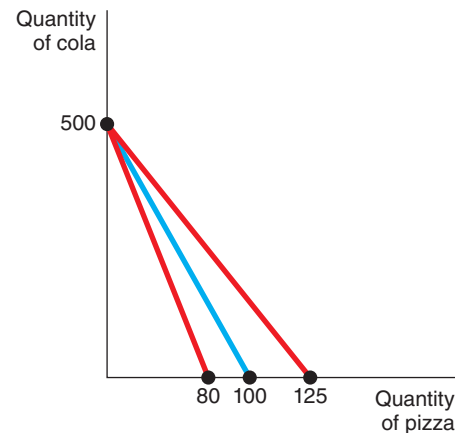
The slope of the budget constraint has now changed. The ratio of the price of cola to the price of pizza is now $\frac{5}{10}$ so the slope of the budget line is -2 . For every 1 litre of cola the consumer acquires he has to give up half a pizza and for every 1 extra pizza bought the consumer now has to sacrifice 2 litres of cola. Notice that as the price of cola has risen the consumer now has to sacrifice fewer litres of cola to purchase every additional pizza, but if the consumer is switching from cola to pizza he has to give up more pizza to buy an additional litre of cola.

If the price of cola were to fall but the price of pizza stayed the same, the budget line would pivot outwards as shown in Figure 5.3. If the price of cola fell to €1.60 the consumer will now be able to afford to buy more cola with his income. If all income was devoted to buying cola the consumer would now be able to purchase 625 litres of cola.

A Change in the Price of Pizza The opposite happens if the price of pizza were to change but the price of cola stayed the same. Assume the price of cola is €2 but the price of pizza rises to €12.50. If the consumer devotes all his income to pizza he can now only afford to buy 80 pizzas with his €1,000 income. The budget constraint would pivot inwards and its slope would change. The ratio of the price of cola to the price of pizza is now $\frac{2}{12.5}$ and the slope is 6.25. The inverse of the slope is 0.16. For every 1 litre of cola the consumer acquires he has to give up 0.16 of a pizza and for every 1 extra pizza bought the consumer now has to sacrifice 6.25 litres of cola. If the price of pizza falls to €8 then the consumer would be able to purchase more pizza (125) with his income and so the budget constraint would pivot outwards as shown in Figure 5.4.

FIGURE 5.4**A Change in the Price of Pizza**

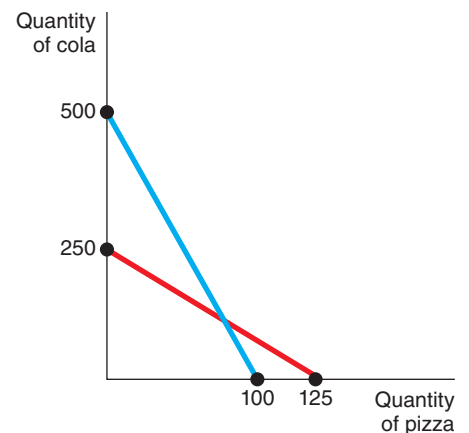
A change in the price of pizza ceteris paribus would cause a pivot in the budget constraint. If the price of pizza fell then the budget constraint would pivot outwards and if the price of pizza rose the budget constraint would pivot inwards.



A change in the Price of Both Goods If the price of both goods changes then the shape of the budget line would depend on the relative change in the prices of the two goods. The slope would still be the ratio of the price of cola to the price of pizza. Figure 5.5 shows a rise in the price of cola from €2 to €4 and a fall in the price of pizza from €10 to €8. If the consumer devoted all his €1,000 per month income to cola he could now afford to buy 250 litres of cola and if he devoted all his income to pizza he could now buy 125 pizzas per month. The ratio of the price of cola to the price of pizza would be $\frac{4}{8}$ and so the slope of the budget constraint would now be 2.

FIGURE 5.5**A Change in the Price of Both Cola and Pizza**

The effect of a change in the price of both goods on the budget constraint depends on the relative change in the prices of the two. In this example the price of cola has risen and the price of pizza has fallen causing the budget constraint to change shape. The slope is now 0.5.



PREFERENCES: WHAT THE CONSUMER WANTS

The budget constraint shows what combination of goods the consumer can afford given his income and the prices of the goods. The consumer's choices also depend on his preferences. We will continue our analysis using cola and pizza as the consumer's **choice set** – the set of alternatives available to the consumer.

choice set the set of alternatives available to the consumer

Representing Preferences with Indifference Curves

The consumer's preferences allow him to choose among different bundles of cola and pizza. The SEM assumes that consumers behave rationally and that, if you offer two different bundles, they chose the bundle that best suits their tastes. Remember, we measure the level of satisfaction in terms of the utility it yields. We can represent consumer preferences in relation to the utility that different bundles of goods provide. If a consumer prefers one bundle of goods to another, the assumption of the SEM is that the first provides more utility than the second. If two bundles yield the same utility then the consumer is said to be *indifferent* between them. We can represent these preferences as indifference curves. An **indifference curve** shows the bundles of consumption that yield the same utility, i.e. makes the consumer equally happy. You can think of an indifference curve as an 'equal-utility' curve.

indifference curve a curve that shows consumption bundles that give the consumer the same level of satisfaction

In our example we are going to use indifference curves that show the combination of cola and pizza with which the consumer is equally satisfied.

The model of preferences which we are looking at has certain assumptions. These laws are based on two axioms (points of reference or starting points).

The Axiom of Comparison Given any two bundles of goods, A and B, representing consumption choices, a consumer can compare these bundles such that A is preferred to B, B is preferred to A or the consumer is indifferent between A and B.

The Axiom of Transitivity Given any three bundles of goods, A, B and C, if the consumer prefers A to B and prefers B to C then he must prefer A to C. Equally, if the consumer is indifferent between A and B and is also indifferent between B and C then they must be indifferent between A and C.

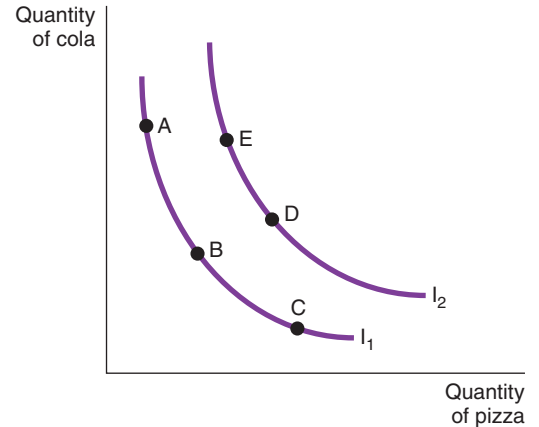
Representing Indifference Curves Graphically

We can represent indifference curves graphically. The quantity of cola is on the vertical axis and the quantity of pizza is on the horizontal axis. The graph is sometimes called an indifference map. The map contains an infinite number of indifference curves. Figure 5.6 shows two of the consumer's many indifference curves.

The points A, B and C on indifference curve I_1 in Figure 5.6 all represent different combinations of cola and pizza. The consumer is indifferent between these combinations. However, indifference curve I_2 is further to the right than curve I_1 and given the assumption that more is preferred to less, the consumer would prefer to be on the highest indifference curve possible. Any point on I_2 , therefore, is preferred to any point on I_1 . Because the consumer prefers points on higher indifference curves, bundles of goods on higher indifference curves provide higher utility.

FIGURE 5.6**The Consumer's Preferences**

The consumer's preferences are represented with indifference curves, which show the combinations of cola and pizza that make the consumer equally satisfied. Because of the assumption that more is preferred to less, points on a higher indifference curve (I_2 here) are preferred to points on a lower indifference curve (I_1). A point along an indifference curve such as point B on indifference curve I_1 , represents a bundle or combination of goods, cola and pizza in this case. The consumer is indifferent between any point along an indifference curve such as A, B or C along indifference curve I_1 . Points D and E on indifference curve I_2 also represent combinations of goods between which the consumer is indifferent but any point on indifference curve I_2 is preferred to any point on indifference curve I_1 .



We can use indifference curves to rank any two bundles of goods. For example, the indifference curves tell us that point D is preferred to point B because point D is on a higher indifference curve than point B. (That conclusion may be obvious, however, because point D offers the consumer both more pizza and more cola.) The indifference curves also tell us that point D is preferred to point A because point D is on a higher indifference curve. Even though point D has less cola than point A, it has more than enough extra pizza to make the consumer prefer it. By seeing which point is on the higher indifference curve, we can use the set of indifference curves to rank any combinations of cola and pizza.

Four Properties of Indifference Curves

Because indifference curves represent a consumer's preferences, they have certain properties that reflect those preferences.

Property 1: Higher indifference curves (further to the upper right) are preferred to lower ones. This is because we assume that consumers prefer more of something to less of it. Higher indifference curves represent larger quantities of goods than lower indifference curves. Thus, the consumer prefers being on higher indifference curves.

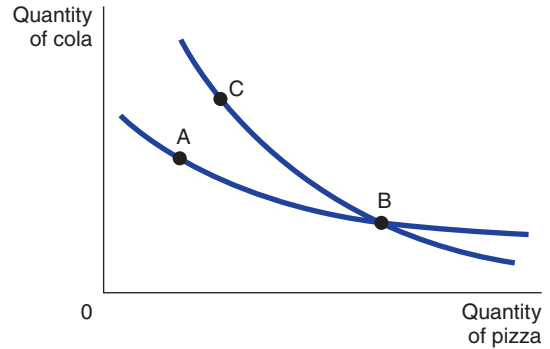
Property 2: Indifference curves are downwards sloping. The slope of an indifference curve reflects the rate at which the consumer is willing to substitute one good for the other. In most cases, the consumer likes both goods. Therefore, if the quantity of one good is reduced, the quantity of the other good must increase in order for the consumer to be equally happy. For this reason, most indifference curves slope downwards.

Property 3: Indifference curves do not cross. To see why this is true, suppose that two indifference curves did cross, as in Figure 5.7. Then, because point A is on the same indifference curve as point B, the two points would make the consumer equally happy. In addition, because point B is on the same indifference curve as point C, these two points would make the consumer equally happy. But these conclusions imply that points A and C would also make the consumer equally happy, even though point C has more of both goods. This contradicts the *axiom of transitivity* and, thus, indifference curves cannot cross.

Property 4: Indifference curves are bowed inward. The slope of an indifference curve is the marginal rate of substitution which we will cover in more detail later. The marginal rate of substitution usually depends on the amount of each good the consumer is currently consuming. In particular, because people are more willing to trade away goods that they have in abundance and less willing to trade away goods of which they have little, the indifference curves are bowed inward. As an example, consider Figure 5.8. At point A, because the consumer has a lot of cola and only a little pizza, he is very hungry but not very thirsty.

FIGURE 5.7**The Impossibility of Intersecting Indifference Curves**

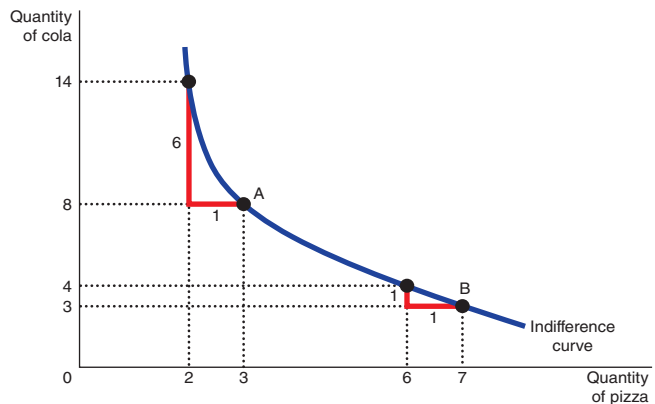
A situation like this can never happen. According to these indifference curves, the consumer would be equally satisfied at points A, B and C, even though point C has more of both goods than point A.



To induce the consumer to give up 1 pizza, the consumer has to be given 6 litres of cola: the marginal rate of substitution is 6 litres per pizza. By contrast, at point B, the consumer has little cola and a lot of pizza, so he is very thirsty but not very hungry. At this point, he would be willing to give up 1 pizza to get 1 litre of cola: the marginal rate of substitution is 1 litre per pizza. Thus, the bowed shape of the indifference curve reflects the consumer's greater willingness to give up a good that he already has in large quantity.

FIGURE 5.8**Bowed Indifference Curves**

Indifference curves are usually bowed inward. This shape occurs because at point A, the consumer has little pizza and much cola, so he requires a lot of extra cola to induce him to give up one of the pizzas: 6 litres of cola per pizza, to be precise. At point B, the consumer has much pizza and little cola, so he requires only a little extra cola to induce him to give up one of the pizzas – 1 litre of cola in this example.

**Total and Marginal Utility**

Remember, we assume that consumers prefer more to less. This does not mean that if a consumer eats more and more pizza or drinks more and more cola that the utility on an additional unit consumed is always the same. We need to distinguish between total utility and marginal utility. **Total utility** is the satisfaction that consumers gain from consuming a product. The **marginal utility** of consumption is the increase in utility that the consumer gets from an additional (marginal) unit of that good.

total utility the satisfaction gained from the consumption of a good

marginal utility the addition to total utility as a result of consuming one extra unit of a good

Imagine that you have been working hard and now realize that you are very hungry. You head to the university canteen and buy a pizza. You eat the first slice of pizza very quickly because you are so hungry.

If you were asked to rate the satisfaction from consuming that first slice (out of 10) you might rate it as 10 out of 10. You then turn to the second slice, that too is good but when you come to rate it you don't give it quite as many out of 10 as the first slice – say you give the second slice 9 out of 10. The total utility of the 2 slices of pizza is 19 but the additional utility of the first slice was 10 and of the second slice, 9. You now consume the third slice. By now your immediate hunger has been satisfied and you find the third slice is not as satisfying, and you rate this as 7 out of 10. The total utility has risen to 26 but the marginal utility of the third slice is 7. As you finish off the remaining three slices you find you did not really enjoy the last slice – you might even have decided to leave part of it as you are now full. If someone now offered to buy you a second pizza you might refuse – eating one extra slice might just result in you being sick – if this was the case you might even rate this next slice as having negative utility.

The tendency for the total utility from consumption to rise but at a slower rate with additional units of consumption is called diminishing marginal utility. **Diminishing marginal utility** refers to the tendency for the additional satisfaction from consuming extra units of a good to fall.

diminishing marginal utility refers to the tendency for the additional satisfaction from consuming extra units of a good to fall

Most goods are assumed to exhibit diminishing marginal utility; the more of the good the consumer already has, the lower the marginal utility provided by an extra unit of that good.

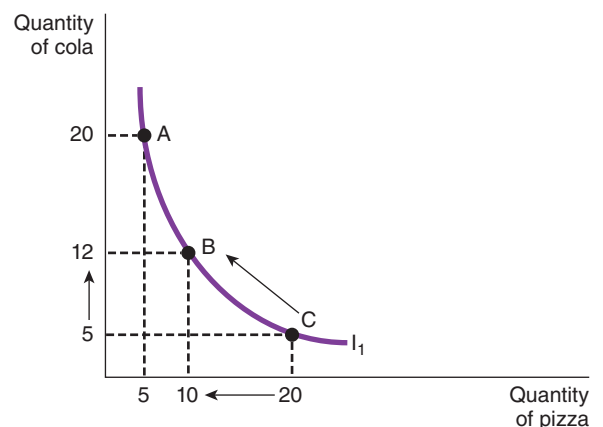
The Marginal Rate of Substitution

Figure 5.9 shows an indifference curve and three combinations of cola and pizza represented by points A, B and C. We know that the consumer is indifferent between these combinations. Assume the consumer starts at the combination of cola and pizza represented by point C. Not surprisingly, if the consumer's consumption of pizza is reduced, say from point C to point B, consumption of cola must increase to keep him equally happy. If consumption of pizza is reduced again, from point B to point A, the amount of cola consumed must increase yet again.

FIGURE 5.9

The Marginal Rate of Substitution

The slope of an indifference curve is not constant throughout its length but changes at every point. The marginal rate of substitution measures that rate at which a consumer is prepared to substitute one good for another. In this example, to acquire an additional 7 units of cola the consumer is willing to give up 10 units of pizza.



Starting at point C, moving to point B the consumer is willing to give up 5 pizzas to increase consumption of cola from 5 litres to 12 litres. We know, however, that as the consumption of cola increases, that consumption will be subject to diminishing marginal utility. In our example, the consumer starts off with a relatively large amount of pizza and a small amount of cola. As we have seen earlier, it is logical to assume

that the consumer would be willing to give up relatively large amounts of pizza to acquire some additional cola. However, moving from point B to point C is a slightly different matter. The situation is almost reversed—to gain additional litres of cola the consumer is willing to sacrifice fewer pizzas.

The rate at which consumers are willing to substitute one good for another is called the marginal rate of substitution. The slope at any point on an indifference curve equals the rate at which the consumer is willing to substitute one good for the other. The **marginal rate of substitution** (MRS) between two goods depends on their marginal utilities. In this case, the marginal rate of substitution measures how much cola the consumer requires in order to be compensated for a one-unit reduction in pizza consumption. For example, if the marginal utility of cola is twice the marginal utility of pizza, then a person would need 2 units of pizza to compensate for losing 1 unit of cola, and the marginal rate of substitution equals -2 . More generally, the marginal rate of substitution (and thus the negative of the slope of the indifference curve) equals the marginal utility of one good divided by the marginal utility of the other good $\left(\frac{MU_x}{MU_y}\right)$.

marginal rate of substitution the rate at which a consumer is willing to trade one good for another

Notice that because the indifference curves are not straight lines, the marginal rate of substitution is not the same at all points on a given indifference curve.

Two Extreme Examples of Indifference Curves

The shape of an indifference curve tells us about the consumer's willingness to trade one good for the other. When the goods are easy to substitute for each other, the indifference curves are less bowed; when the goods are hard to substitute, the indifference curves are very bowed. To see why this is true, let's consider the extreme cases.

Perfect Substitutes Suppose that someone offered you bundles of 50 cent coins and 10 cent coins. How would you rank the different bundles?

Most probably, you would care only about the total monetary value of each bundle. If so, you would judge a bundle based on the number of 50 cent coins plus five times the number of 10 cent coins. In other words, you would always be willing to trade one 50 cent coin for five 10 cent coins, regardless of the number of coins in either bundle. Your marginal rate of substitution between 10 cent coins and 50 cent coins would be a fixed number: 5.

We can represent your preferences over 50 cent coins and 10 cent coins with the indifference curves in panel (a) of Figure 5.10. Because the marginal rate of substitution is constant, the indifference curves are straight lines. In this extreme case of straight indifference curves, we say that the two goods are **perfect substitutes**.

perfect substitutes two goods with straight line indifference curves

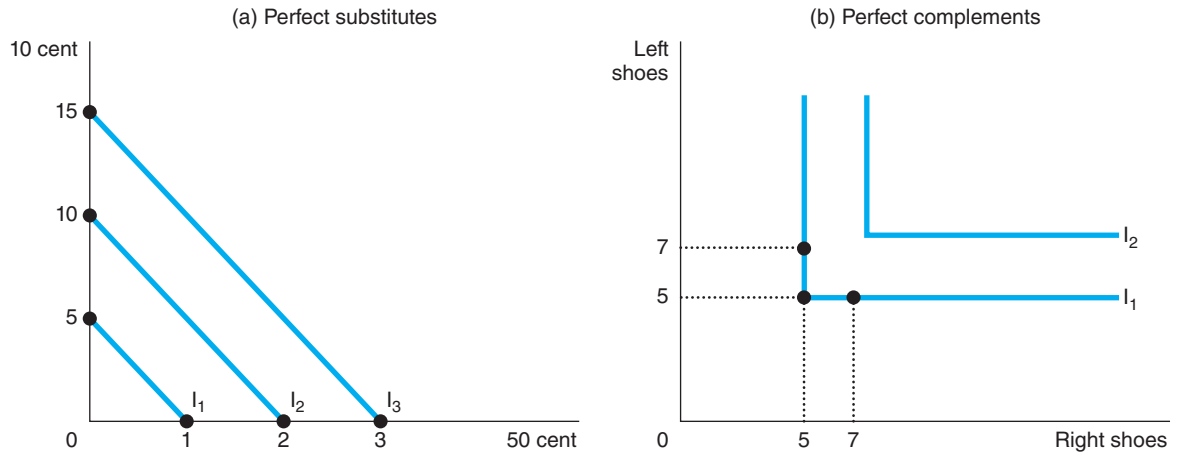
Perfect Complements Suppose now that someone offered you bundles of shoes. Some of the shoes fit your left foot, others your right foot. How would you rank these different bundles?

In this case, you might care only about the number of pairs of shoes. In other words, you would judge a bundle based on the number of pairs you could assemble from it. A bundle of 5 left shoes and 7 right shoes yields only 5 pairs. Getting 1 more right shoe has no value if there is no left shoe to go with it.

We can represent your preferences for right and left shoes with the indifference curves in panel (b) of Figure 5.10. In this case, a bundle with 5 left shoes and 5 right shoes is just as good as a bundle with 5 left shoes and 7 right shoes. It is also just as good as a bundle with 7 left shoes and 5 right shoes.

FIGURE 5.10**Perfect Substitutes and Perfect Complements**

When two goods are easily substitutable, such as 50 cent and 10 cent coins, the indifference curves are straight lines, as shown in panel (a). When two goods are strongly complementary, such as left shoes and right shoes, the indifference curves are right angles, as shown in panel (b).



The indifference curves, therefore, are right angles. In this extreme case of right-angle indifference curves, we say that the two goods are **perfect complements**.

perfect complements two goods with right-angle indifference curves

In the real world, of course, most goods are neither perfect substitutes (like coins of different denominations) nor perfect complements (like right shoes and left shoes). More typically, the indifference curves are bowed inward, but not so bowed as to become right angles.

SELF TEST Why is the amount of a good a person is consuming at a given point in time an important factor in determining the marginal rate of substitution? What happens to the total and marginal utility of two goods, x and y , if a consumer has a large quantity of good x but hardly any of good y and then opts to consume more of good x ?

OPTIMIZATION: WHAT THE CONSUMER CHOOSES

One of the working assumptions of the SEM is that the consumer will seek to maximize utility subject to the constraint of a limited income. This is an example of a constrained optimization problem. How is this constrained optimization problem solved?

The Consumer's Optimal Choices

Using our cola and pizza example, the consumer would like to end up with the best possible combination of cola and pizza – that is, the combination on the highest possible indifference curve. But the consumer is constrained by the resources available to him – his income as shown by the budget constraint.

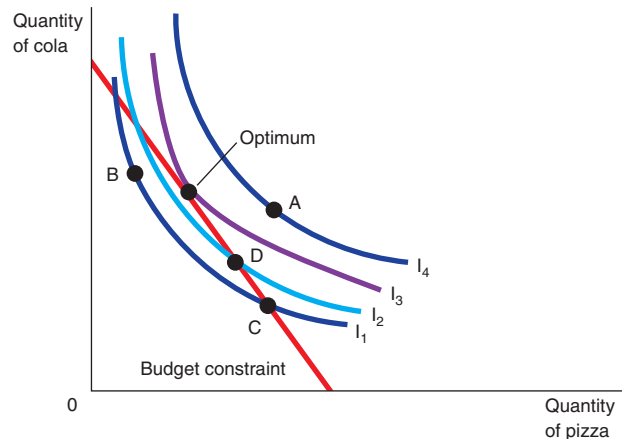
Figure 5.11 shows the consumer's budget constraint and three of his many indifference curves. The highest indifference curve that the consumer can reach (I_3 in the figure) is the one that just barely touches the budget constraint. The point at which this indifference curve and the budget constraint touch is called the *optimum*. The consumer would prefer point A, but he cannot afford that point because it lies above his budget constraint. The consumer can afford point B, but that point is on a lower indifference curve and, therefore, provides the consumer less satisfaction. The combination of cola and pizza at point C is affordable being just on the budget constraint, but the consumer is not in equilibrium because there is an incentive for him to change his consumption choices and reach a higher indifference curve. What this means is that the consumer can reallocate his spending decisions and get more utility from his limited income. There is an incentive for him to reduce consumption of pizza and increase consumption of cola at point D on indifference curve I_2 . By doing this the consumer is getting more utility from an additional cent spent on more cola compared to the marginal utility spent on another pizza. This is entirely logical. If you could spend one cent on more cola and get an extra 7 utils of utility compared to an extra 5 utils you would get from more pizza with the same cent, it makes sense to buy more cola.

However, this is still not the optimum because the consumer can continue to reallocate his spending decisions reaching ever higher indifference curves (remember there are an infinite amount on the map) until the marginal utility of the last cent spent on cola is equal to the marginal utility of the last cent spent on pizza. The optimum represents the best combination of consumption of cola and pizza available to the consumer.

FIGURE 5.11

The Consumer's Optimum

The consumer chooses the point on his budget constraint that lies on the highest achievable indifference curve. At this point, called the optimum, the marginal rate of substitution equals the relative price of the two goods. Here the highest indifference curve the consumer can reach is I_3 . The consumer prefers point A, which lies on indifference curve I_4 , but the consumer cannot afford this bundle of cola and pizza. In contrast, point B is affordable, but because it lies on a lower indifference curve, the consumer does not prefer it.



At the point of consumer equilibrium, the optimum, the slope of the indifference curve equals the slope of the budget constraint. We say that the indifference curve is *tangential* to the budget constraint. The slope of the indifference curve is the marginal rate of substitution between cola and pizza, and the slope of the budget constraint is the relative price of cola and pizza. Thus, the consumer chooses consumption of the two goods so that the marginal rate of substitution equals the relative price.

At the consumer optimum, the marginal rate of substitution equals the ratio of prices. That is:

$$MRS = \frac{P_x}{P_y}$$

Because the marginal rate of substitution equals the ratio of marginal utilities, we can write this condition for optimization as:

$$\frac{MU_x}{MU_y} = \frac{P_x}{P_y}$$

Now rearrange this expression to become:

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$$

This equation has a simple interpretation: at the optimum, the marginal utility per euro spent on good x equals the marginal utility per euro spent on good y . At any other point, the consumer is not in equilibrium. Why? As we saw above, this is because the consumer could increase utility by changing behaviour, switching spending from the good that provided lower marginal utility per euro and more on the good that provided higher marginal utility per euro.

At the consumer's optimum, the consumer's valuation of the two goods (as measured by the marginal rate of substitution) equals the market's valuation (as measured by the relative price). As a result of this consumer optimization, market prices of different goods reflect the value that consumers place on those goods.

CASE STUDY Christmas and the Rational Consumer

At the time of writing, it is a few weeks before the Christmas holiday. Millions of people throughout Europe will either travel to shopping centres or go online to buy gifts for friends and relatives. The media are full of adverts trying to persuade people to buy particular products and TV adverts are dominated by food, drink, chocolate, perfumes, toys and electrical goods.

Does our theory of consumer choice provide any insights into consumer behaviour at this time of the year? Do consumers act rationally to allocate their limited income? Many people would like to think that they do; most appreciate that they have a limited income and when economic times are hard (which they are for many countries at the time of writing) there is even more of an incentive to make budgets stretch that bit further and give value for money.

There is evidence, however, that consumers may not behave as rationally as they might like to think. The first problem is that when buying gifts consumers are not the end user and so are anticipating the utility of a third person – in other words, will the gift receiver actually like the present when they open it? The next problem is that consumers are influenced in their consumption choices by the information they are bombarded with by retailers – both online and 'bricks and mortar'.

As consumers walk into shops they are often met with 'offers' – the number of people who stop and buy something as a result of these 'offers' is striking. In some cases they might be for so-called 'big ticket items' such as TVs which may be deliberately sold below cost to encourage purchase. Where is the rationality in that for the retailer? One reason is that in buying the TV the consumer invariably then also purchases other accessories which they need – cables, stands, wall mounting brackets, maybe even a games console or some DVDs to complement the wonderful new TV and before you know it the amount spent is far above any savings gained from purchasing the TV. We might also ask whether the consumer actually needed the TV in the first place – is this what they set out to buy? In some cases, of course, it is, but in others it is an impulse purchase and can that, by definition, be rational?

Doing the Christmas grocery shopping might be another case where rationality is compromised. Many stores provide information on their receipt or online order form telling consumers how much they have 'saved' as a result of their purchase decisions. The consumer might have decided to buy a six-pack of wine for €48 which will 'save' them €6 – did they intend to buy a six-pack in the first place? If not then they may end up spending a great deal more than they originally intended. It doesn't matter of course because 'it will all get drunk at some point so I might as well buy it!' Is that a rational statement?

The 'savings' list might sound impressive but the store does not offset the savings they tell you about by allocating the overheads of shopping – the time spent queuing, the cost of the petrol and vehicle depreciation used to get to the store, the parking fee, what the shopper could have been doing if they were not walking round a supermarket or sitting in front of their computer making their consumption choice.

Then there is the cheerful retail operative who proudly announces to you that as a result of buying the two-for-one shirt offer you have 'saved €30'. The answer? No, I haven't, I have spent €30! I didn't need two

(Continued)

shirts (I probably did not *need* one shirt) and if I was being asked to hand over €60 for the item regardless of how many shirts, I may not have made the decision to spend that much money. If I was never going to spend it then it is not a saving. Many consumers, however, do tell their friends that they 'saved' lots as a result of their trip, despite these facts. Psychologists call it *confirmation bias*.

There is also research which shows that far from being rational in our decision making when shopping, we can be deeply irrational. Our brains react differently to information we see when we are shopping and retailers know this. On the one hand we are primed when we go shopping. Psychologists have shown that when humans are given information which creates an association they are more likely to behave in ways which confirm the association. If the brain is primed to look for bargains or offers we are more likely to make a purchase decision when confronted with information which seems to represent a 'bargain'. If prices are lower *than anticipated* then the decision making part of the brain registers this and we are more likely to make the purchase as a result. Equally, if prices are much higher than anticipated a different part of the brain associated with registering pain is activated and we avoid the purchase.

Overall, therefore, whilst our model of consumer behaviour provides some insights, as with any model it is a representation of reality and not reality itself. There is now a great deal of research being conducted which is shedding more light on consumer behaviour which casts doubt on just how rational we are capable of being.



Many stores provide information on their receipt or online order form telling consumers how much they have 'saved' as a result of their purchase decisions

How Changes in Income Affect the Consumer's Choices

Suppose that income increases. We know that if income increases there is a parallel shift of the budget constraint with the same slope of the initial budget line because the relative price of the two goods has not changed.

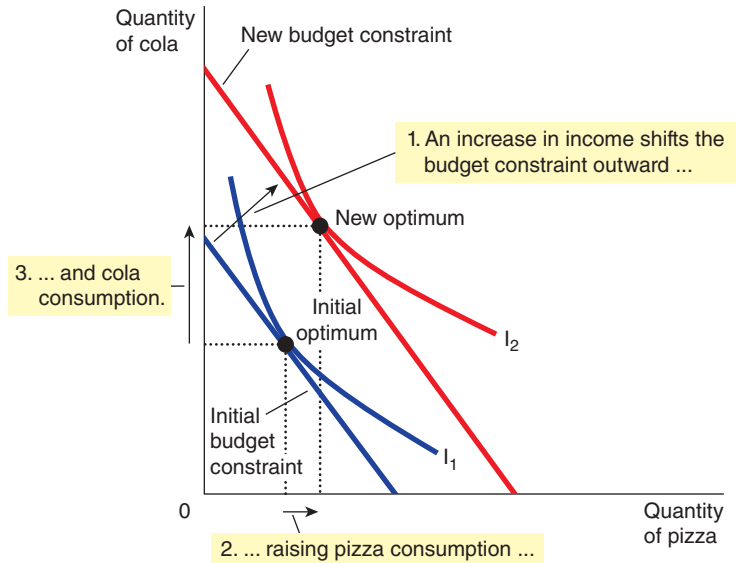
The increase in income means there is an incentive for the consumer to reallocate his spending decisions to increase utility and choose a better combination of cola and pizza. In other words, the consumer can now reach a higher indifference curve as shown in Figure 5.12. The consumer reallocates income until he reaches a new optimum labelled 'new optimum'.

Notice that in Figure 5.12 the consumer chooses to consume more cola and more pizza. Although the logic of the model does not require increased consumption of both goods in response to increased income, this situation is the most common one. Remember, if a consumer wants more of a good when his income rises, economists call it a normal good. The indifference curves in Figure 5.12 are drawn under the assumption that both cola and pizza are normal goods.

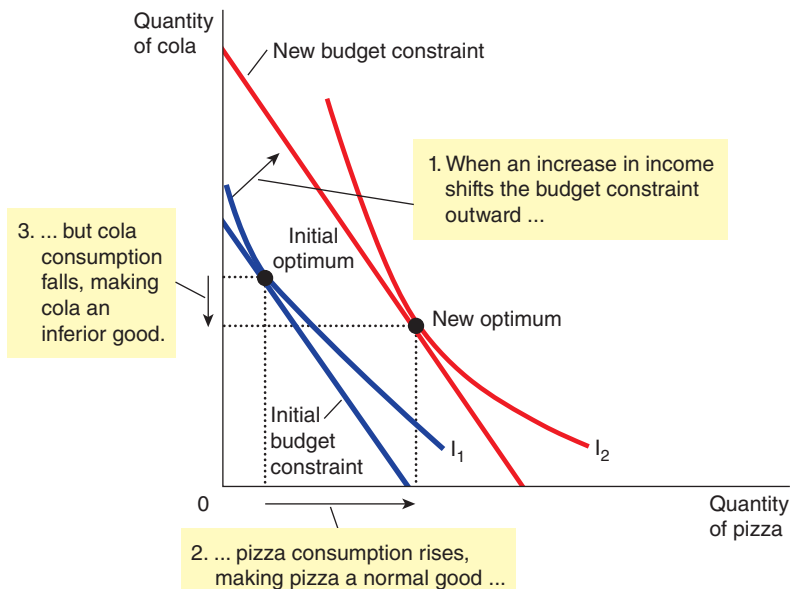
Figure 5.13 shows an example in which an increase in income induces the consumer to buy more pizza but less cola. If a consumer buys less of a good when his income rises, economists call it an inferior good. Figure 5.13 is drawn under the assumption that pizza is a normal good and cola is an inferior good.

FIGURE 5.12**An Increase in Income**

When the consumer's income rises, the budget constraint shifts out. If both goods are normal goods, the consumer responds to the increase in income by buying more of both of them. Here the consumer buys more pizza and more cola.

**FIGURE 5.13****An Inferior Good**

A good is an inferior good if the consumer buys less of it when his income rises. Here cola is an inferior good: when the consumer's income increases and the budget constraint shifts outward, the consumer buys more pizza but less cola.



Although most goods are normal goods, there are some inferior goods in the world. One example is bus rides. High-income consumers are more likely to own cars and less likely to ride the bus than low-income consumers. Bus rides, therefore, are an inferior good.

How Changes in Prices Affect the Consumer's Choices

Suppose that the price of cola falls from €2 to €1 a litre. We have seen how a change in the price of any good causes the budget constraint to pivot. With his available income of €1,000 the consumer can now

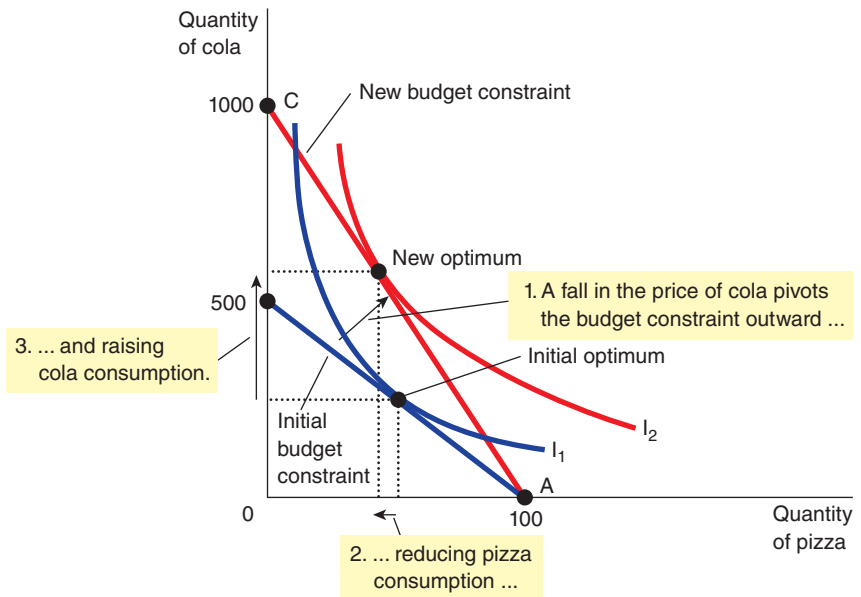
buy twice as many litres of cola than before but the same amount of pizza. Figure 5.14 shows that point A stays the same (100 pizzas). Yet if the consumer spends his entire income of €1,000 on cola, he can now buy 1,000 rather than only 500 litres. Thus, the end point of the budget constraint pivots outwards from point B to point C.

The pivoting of the budget constraint changes its slope. Because the price of cola has fallen to €1 from €2, while the price of pizza has remained €10, the consumer can now trade a pizza for 10 rather than 5 litres of cola. As a result, the new budget constraint is more steeply sloped.

FIGURE 5.14

A Change in Price

When the price of cola falls, the consumer's budget constraint pivots outward and changes slope. The consumer moves from the initial optimum to the new optimum, which changes his purchases of both cola and pizza. In this case, the quantity of cola consumed rises and the quantity of pizza consumed falls.



How such a change in the budget constraint alters the consumption of both goods depends on the consumer's preferences. For the indifference curves drawn in Figure 5.14, the consumer buys more cola and less pizza.

Income and Substitution Effects

In Chapter 3 we took a brief look at the **income effect** and the **substitution effect** as reasons why a fall in price leads to a rise in quantity demanded.

income effect the change in consumption that results when a price change moves the consumer to a higher or lower indifference curve

substitution effect the change in consumption that results when a price change moves the consumer along a given indifference curve to a point with a new marginal rate of substitution

Consider this thought experiment: If the price of cola falls you are now able to purchase more cola with your income – you are *in effect* richer. Because you are richer, you can buy both more cola and more pizza. For example, assume your income is €1,000 and the initial price of cola is €2 and pizza is €10. You currently

buy 250 litres of cola and 50 pizzas. If the price of cola falls to €1 per litre you can adjust your spending and now buy 300 litres of cola (spending €300) and use the remaining €700 to buy more pizza (70 pizzas). This is the income effect.

Second, note that now that the price of cola has fallen, you get more litres of cola for every pizza that you give up. Because pizza is now *relatively* more expensive, you might decide to buy less pizza and more cola. This is the substitution effect.

Both of these effects occur when prices change. The decrease in the price of cola makes the consumer better off. If cola and pizza are both normal goods, the consumer will want to spread this improvement in his purchasing power over both goods. This income effect tends to make the consumer buy more pizza and more cola. Yet, at the same time, consumption of cola has become less expensive relative to consumption of pizza. This substitution effect tends to make the consumer choose more cola and less pizza.

The end result of these two effects is that the consumer certainly buys more cola, because the income and substitution effects both act to raise purchases of cola. But it is ambiguous whether the consumer buys more pizza, because the income and substitution effects work in opposite directions. This conclusion is summarized in Table 5.1.

TABLE 5.1**Income and Substitution Effects When the Price of Cola Falls**

| Good | Income effect | Substitution effect | Total effect |
|-------|--|--|--|
| Cola | Consumer is richer, so he buys more cola. | Cola is relatively cheaper, so consumer buys more cola. | Income and substitution effects act in same direction, so consumer buys more cola. |
| Pizza | Consumer is richer, so he buys more pizza. | Pizza is relatively more expensive, so consumer buys less pizza. | Income and substitution effects act in opposite directions, so the total effect on pizza consumption is ambiguous. |

We can interpret the income and substitution effects using indifference curves:

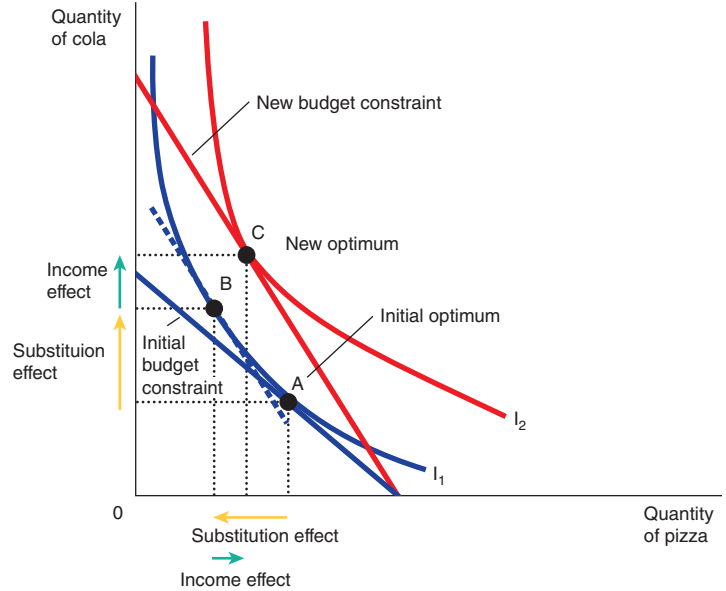
- The income effect is the change in consumption that results from the movement to a higher indifference curve.
- The substitution effect is the change in consumption that results from being at a point on an indifference curve with a different marginal rate of substitution.

Figure 5.15 shows graphically how to decompose the change in the consumer's decision into the income effect and the substitution effect. When the price of cola falls, the consumer moves from the initial optimum, point A, to the new optimum, point C. We can view this change as occurring in two steps. First, the consumer moves *along* the initial indifference curve I_1 from point A to point B – this is the substitution effect. The consumer is equally happy at these two points, but at point B the marginal rate of substitution reflects the new relative price. (The dashed line through point B reflects the new relative price by being parallel to the new budget constraint.) Next, the consumer *shifts* to the higher indifference curve I_2 by moving from point B to point C – this is the income effect. Even though point B and point C are on different indifference curves, they have the same marginal rate of substitution. That is, the slope of the indifference curve I_1 at point B equals the slope of the indifference curve I_2 at point C.

Although the consumer never actually chooses point B, this hypothetical point is useful to clarify the two effects that determine the consumer's decision. Notice that the change from point A to point B represents a pure change in the marginal rate of substitution without any change in the consumer's welfare. Similarly, the change from point B to point C represents a pure change in welfare without any change in the marginal rate of substitution. Thus, the movement from A to B shows the substitution effect, and the movement from B to C shows the income effect.

FIGURE 5.15**Income and Substitution Effects**

The effect of a change in price can be broken down into an income effect and a substitution effect. The substitution effect – the movement along an indifference curve to a point with a different marginal rate of substitution – is shown here as the change from point A to point B along indifference curve I_1 . The income effect – the shift to a higher indifference curve – is shown here as the change from point B on indifference curve I_1 to point C on indifference curve I_2 .



SELF TEST Draw a budget constraint and indifference curves for cola and pizza. Show what happens to the budget constraint and the consumer's optimum when the price of pizza rises. In your diagram, decompose the change into an income effect and a substitution effect.

Deriving the Demand Curve

Using the logic we have developed so far, we can now look at how the demand curve is derived. The demand curve shows the quantity demanded of a good for any given price. We can view a consumer's demand curve as a summary of the optimal decisions that arise from his budget constraint and indifference curves.

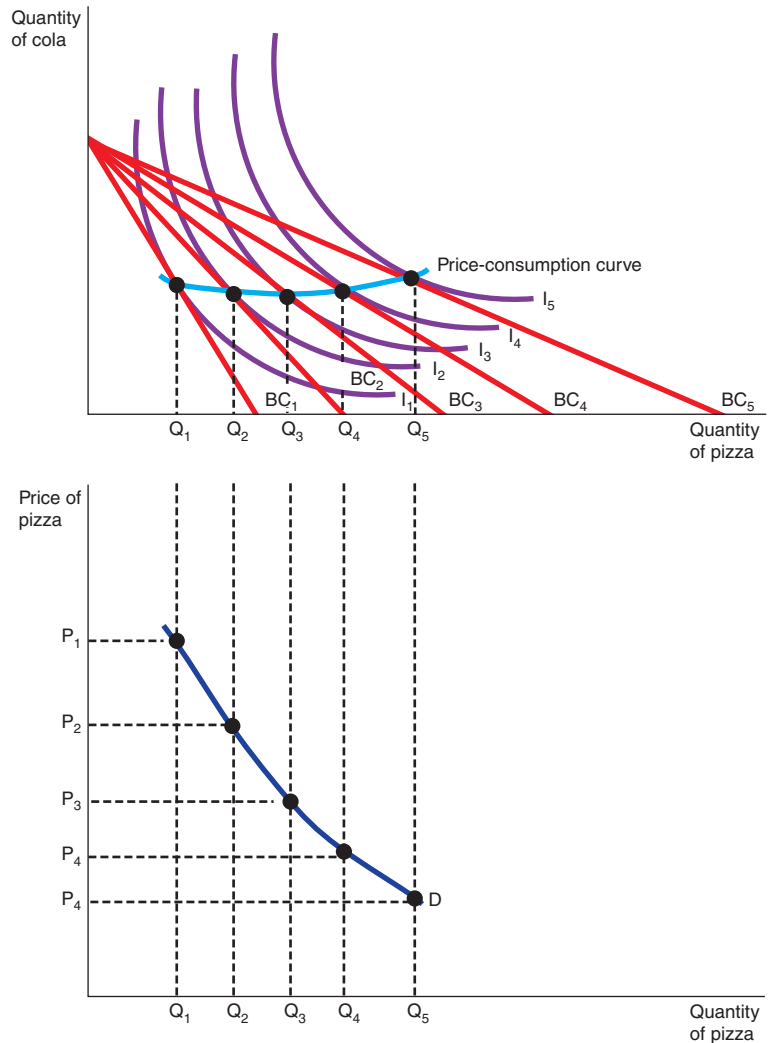
Figure 5.16 considers the demand for pizza. Assume the price of pizza is €10 indicated by the budget constraint BC_1 and a consumer optimum with indifference curve I_1 giving a quantity of pizza bought as Q_1 . A series of other budget constraints labelled BC_2 to BC_5 indicate successive lower prices of pizza and the different consumer optimums indicated by the varying quantities of pizza purchased. The consumer optimums associated with each price of pizza are shown as the **price-consumption curve**. The price-consumption curve shows the consumer optimum for two goods as the price of one of the goods changes *ceteris paribus*.

price-consumption curve a line showing the consumer optimum for two goods as the price of one of the goods changes, assuming incomes and the price of the good are held constant

Figure 5.16 represents the relationship between the change in the price of pizza and the quantity demanded. The price–quantity relationship is plotted on the lower graph to give the familiar demand curve. As the price of pizza falls the quantity demanded rises – the reasons being partly due to the income effect and partly to the substitution effect. The theory of consumer choice, therefore, provides the theoretical foundation for the consumer's demand curve, which we first introduced in Chapter 3.

FIGURE 5.16**Deriving the Demand Curve**

The upper graph shows that when the price of cola falls the consumer's optimum changes and these changes are shown as the price-consumption curve. The demand curve in the lower graph reflects the relationship between the price of pizza and the quantity demanded.



Do All Demand Curves Slope Downwards?

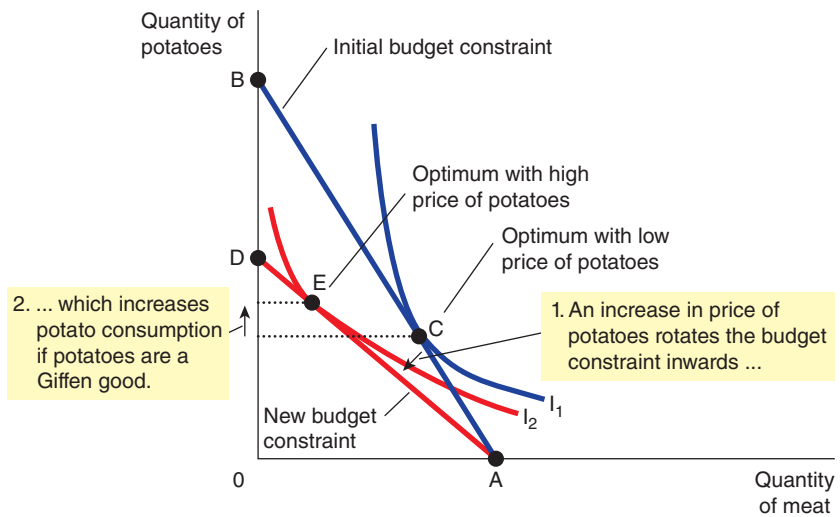
The *law of demand* implies that when the price of a good rises, people buy less of it. This law is reflected in the downwards slope of the demand curve.

As a matter of economic theory, however, demand curves can sometimes slope upwards. In other words, consumers can sometimes violate the law of demand and buy *more* of a good when the price rises. To see how this can happen, consider Figure 5.17. In this example, the consumer buys two goods – meat and potatoes. Initially, the consumer's budget constraint is the line from point A to point B. The optimum is point C. When the price of potatoes rises, the budget constraint shifts inwards and is now the line from point A to point D. The optimum is now point E. Notice that a rise in the price of potatoes has led the consumer to buy a larger quantity of potatoes.

Why is the consumer responding in a seemingly perverse way? The reason is that potatoes here are a strongly inferior good. When the price of potatoes rises, the consumer is poorer. The income effect makes the consumer want to buy less meat and more potatoes. At the same time, because the potatoes have become more expensive relative to meat, the substitution effect makes the consumer want to buy more meat and less potatoes. In this particular case, however, the income effect is so strong that it exceeds the substitution effect. In the end, the consumer responds to the higher price of potatoes by buying less meat and more potatoes.

FIGURE 5.17**A Giffen Good**

In this example, when the price of potatoes rises, the consumer's optimum shifts from point C to point E. In this case, the consumer responds to a higher price of potatoes by buying less meat and more potatoes.



Economists use the term **Giffen good** to describe a good that violates the law of demand. (The term is named after the British economist Robert Giffen, who first noted this possibility.) In this example, potatoes are a Giffen good. Giffen goods are inferior goods for which the income effect dominates the substitution effect. Therefore, they have demand curves that slope upward.

Giffen good a good for which an increase in the price raises the quantity demanded

Economists disagree about whether any Giffen good has ever been discovered. Some historians suggest that potatoes were in fact a Giffen good during the Irish potato famine of the 19th century. Potatoes were such a large part of people's diet (historians estimate that the average working man might eat up to 14 pounds of potatoes a day) that when the price of potatoes rose, it had a large income effect. People responded to their reduced living standard by cutting back on the luxury of meat and buying more of the staple food of potatoes. Thus, it is argued that a higher price of potatoes actually raised the quantity of potatoes demanded.

Whether or not this historical account is true, it is safe to say that Giffen goods are very rare. Some economists (for example, Dwyer and Lindsey, 1984 and Rosen, 1999) have claimed that a legend has built up around Robert Giffen and that the evidence does not support his idea. Others have suggested that rice and wheat in parts of China might exhibit Giffen qualities. The theory of consumer choice does allow demand curves to slope upwards. Yet such occurrences are so unusual that the law of demand is as reliable a law as any in economics (see: Dwyer, G.P. and Lindsay, C.M. (1984) 'Robert Giffen and the Irish Potato'. In *The American Economic Review*, **74**:188–192; Jensen, R and Miller, N. (2008) 'Giffen Behavior and Subsistence Consumption'. In *American Economic Review*, **97**:1553–1577; Rosen, S. (1999) 'Potato Paradoxes'. In *The Journal of Political Economy*, **107**:294–213).

The Income Expansion Path

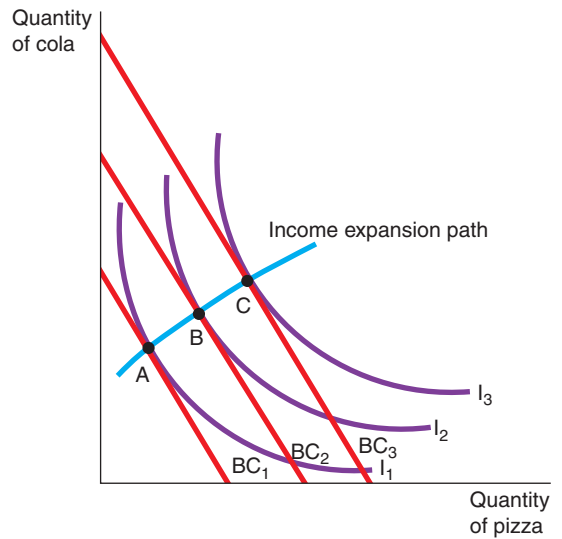
Having looked at conclusions that can be drawn from assuming a change in price (holding all other things constant) we now turn our attention to what happens if we change income (*ceteris paribus*).

How does the rational consumer respond to a change in income? In Chapter 3 we noted that, for a normal good, a rise in income is associated with a rise in demand but, for an inferior good, a rise in income means a fall in demand. We now have the analytical tools to understand why this is the case.

Normal Goods Figure 5.18 shows a series of increases in income represented by three budget constraints BC_1 , BC_2 and BC_3 for cola and pizza. The consumer optimums in each case are indicated by points A, B and C.

FIGURE 5.18**The Income Expansion Path**

As income increases, shown by the shifts in the budget constraint from BC_1 to BC_3 , the consumer optimum changes as shown by the income expansion path. In this example, both cola and pizza are normal goods – as income rises the demand for cola and pizza rises.

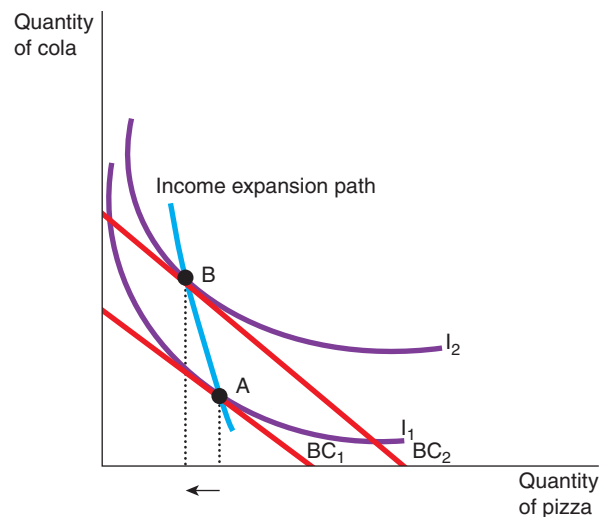


If we connect these points we get the income expansion path which reflects the response of a rational consumer to a change in income. In this example, the increase in income has led to an increase in the consumption of both pizza and cola, and as a result we can conclude that both goods are normal goods. In both goods the income effect outweighs the substitution effect.

Where Pizza is an Inferior Good Figure 5.19 shows a situation where as a result of the increase in income, represented by a shift of the budget constraint from BC_1 to BC_2 , there is a change in the consumer optimum from point A to point B. The income expansion path indicates that as income rises, demand for cola increases (it is a normal good) but the demand for pizza has decreased indicating that it is an inferior good. In this case the substitution effect on pizza of the rise in income has outweighed the income effect.

FIGURE 5.19**The Income Expansion Path Where Pizza is an Inferior Good**

The increase in income has caused the consumer optimum to change from point A to point B. The demand for cola has risen but the demand for pizza has fallen indicating that pizza is an inferior good in this example.

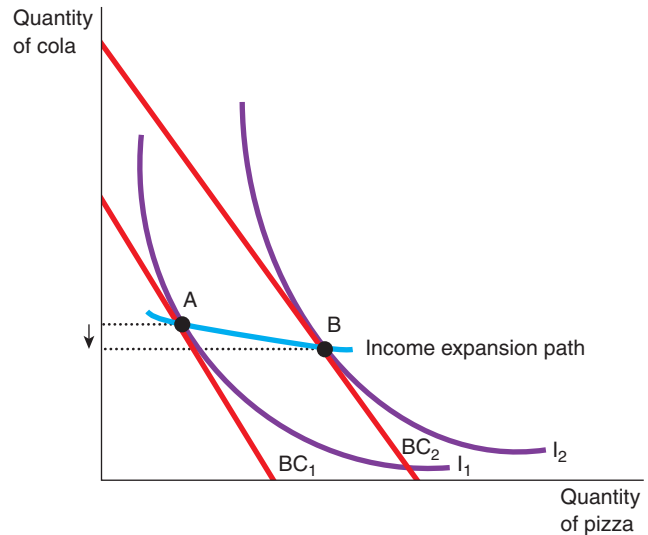


Where Cola is an Inferior Good Figure 5.20 shows a situation again where, as a result of the increase in income, the budget constraint shifts from BC_1 to BC_2 , and the consumer optimum is represented by points A and B. In this case the income expansion path shows that as income rises, demand for cola decreases, showing that it is an inferior good. The demand for pizza has increased as a result of the increase in income indicating that it is a normal good. In this case the substitution effect of the rise in income on cola is greater than the income effect.

FIGURE 5.20

The Income Expansion Path Where Cola is an Inferior Good

The increase in income has caused the consumer optimum to change from point A to point B. The demand for cola has fallen but the demand for pizza has risen indicating that cola is an inferior good in this example.



The Engel Curve

The income expansion path allows us to see an interesting discovery made by German statistician, Ernst Engel (1821–1896). Engel spent some time investigating the relationship between changes in income and spending on broad categories of goods such as food. In 1857, Engel proposed a theory referred to as Engel's law. Engel observed that as income rises the proportion of income spent on food decreases, whereas the proportion of income devoted to other goods, such as leisure, increases. For example, imagine a family of four has a combined annual income of €45,000 and spends €15,000 of that income on food. This represents a third of the family's income spent on food. If the combined income then doubled to €90,000 it is unlikely that spending on food will rise to €30,000; it might rise to €20,000 and if it did then the proportion spent on food would now be just over 22 per cent. Engel's findings have been observed and supported many times since his discovery and have important implications for government policy and for businesses. For example, if incomes are rising, firms selling food will not see their revenues rising in proportion to the change in incomes. In economies where per capita income is very low, any increase in incomes may see higher proportions initially spent on food but then start to decline as these economies change to become emerging economies. Equally, in many countries the poor will spend a higher proportion of their income on food than will those on middle and high incomes. Businesses selling other goods such as those in the leisure industry may find that as incomes rise expenditure on their goods and services increases.

Note that as incomes increase, families may spend a small proportion of their income on food but that does not mean to say that food is not a normal good. At low levels of income, spending on food is important – families have to live. As income increases spending on food increases but the rate at which it increases starts to fall. The income elasticity of demand for food is likely to become more inelastic as income increases. Conversely, the income elasticity of demand for other goods (which we shall call luxuries) increases as income increases and at a faster rate – the income elasticity of demand for luxuries becomes more elastic as income rises.

The upper graph in Figure 5.21 shows two goods: food on the vertical axis and luxuries on the horizontal axis. As income increases, as shown by the three budget constraints BC_1 to BC_3 , the change in consumer

optimum is shown by the three points A, B and C. As income increases, the demand for food and luxuries both increase – both are normal goods. However, the amount of food demanded is increasing at a diminishing rate whereas the increase in the demand for luxuries is greater as income increases. The implication, therefore, is that the demand for food is income inelastic whereas the demand for luxuries is income elastic.

The relationship between income and the demand for luxuries is plotted on the lower graph in Figure 5.21 showing how the demand for luxuries increases as income increases. The line joining points A, B and C is the **Engel curve**. The Engel curve is a line showing the relationship between demand and levels of income.

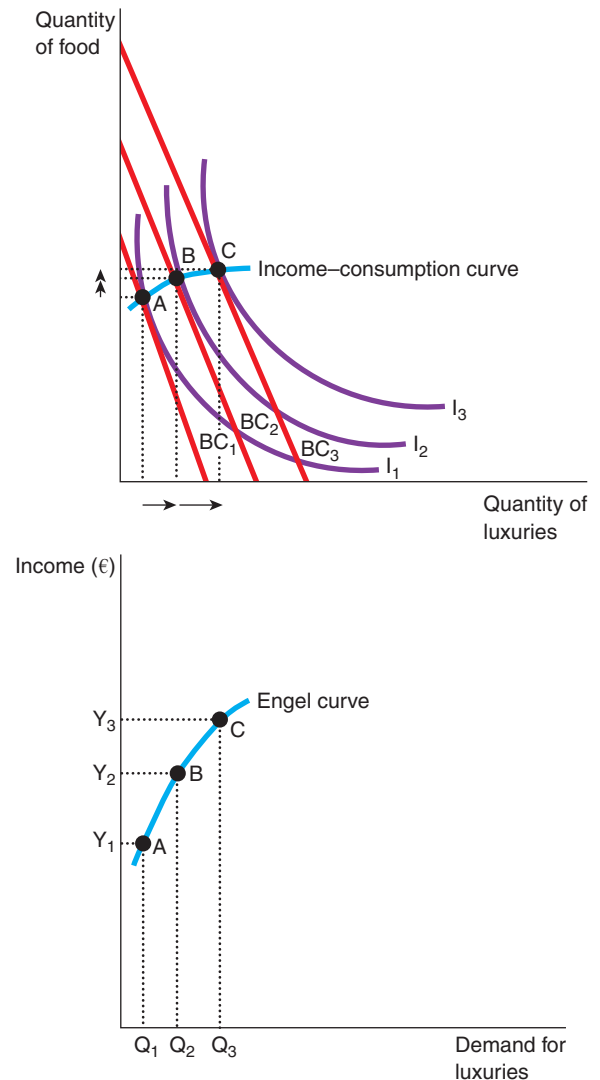
Engel curve a line showing the relationship between demand and levels of income

In this example the increase in demand for luxuries is greater between income levels Y_2 and Y_3 than it was between Y_1 and Y_2 suggesting that the income elasticity of demand is getting greater as income rises. However, we can see that the proportionate increase in quantity demanded of luxuries between Y_1 and Y_2 is less than the proportionate increase in income, which suggests that the demand for luxuries is income inelastic between these income levels.

FIGURE 5.21

The Engel Curve

The upper graph shows the income-consumption curve for two goods, food and luxuries, which are both normal goods. The lower graph plots the change in demand for luxuries against changes in income.



SUMMARY: DO PEOPLE REALLY THINK THIS WAY?

The standard economic model describes how people make decisions based on certain assumptions. It has broad applicability and can explain how a person chooses between cola and pizza, food and leisure, and so on.

At this point, however, you might be tempted to treat the theory of consumer choice with some scepticism. After all, you are a consumer. You decide what to buy every time you walk into a shop. And you know that you do not decide by writing down budget constraints and indifference curves.

It is important to remember that the standard economic model does not try to present a literal account of how people make decisions. It is a model, and, as we first discussed in Chapter 2, models are not intended to be completely realistic. The model does have some merits; consumers are aware that their choices are constrained by their financial resources. Many, given those constraints, will do the best they can to achieve the highest level of satisfaction.

There is, however, considerable evidence that the SEM has limitations in explaining how consumers actually behave. We might like to think we behave rationally but research has shown that our ability to make judgements and decisions can be subject to systematic and consistent flaws – biases – that mean consumer behaviour as represented by the SEM is, at the very best, a limited model of consumer behaviour. Much of the research on this area has been inspired by the work of two psychologists, Daniel Kahneman and Amos Tversky. Indeed, such has been their impact that Kahneman was awarded the Nobel Prize for Economics in 2002 (Tversky sadly died at the relatively young age of 59 in 1996). We now present a brief overview of behavioural approaches to consumer behaviour.

BEHAVIOURAL APPROACHES TO CONSUMER BEHAVIOUR

Doubts about the validity of the SEM have arisen because of a number of reasons but mostly because of doubts about whether consumers really do behave rationally, which is one of the fundamental bases of the SEM. Many of the things we do in life and the decisions we make cannot be explained as those of rational beings. (Rational beings are sometimes referred to by economists as *homoeconomicus*.)

Rather than being rational, humans can be forgetful, impulsive, confused, emotional and short-sighted. Economists have suggested that humans are only ‘near rational’ or that they exhibit ‘bounded rationality’. **Bounded rationality** is the idea that humans make decisions under the constraints of limited, and sometimes unreliable, information, that they face limits to the amount of information they can process and that they face time constraints in making decisions.

bounded rationality the idea that humans make decisions under the constraints of limited, and sometimes unreliable, information

The SEM has an implied assumption that to make a rational decision which maximizes utility, consumers can know everything about the consumption decision that they make and can process all this information very quickly. Research has suggested that this is far from the case. Humans make systematic and consistent mistakes in decision making. We will now outline some of the main errors in judgement and decision making.

People are Overconfident

Imagine that you were asked some numerical questions, such as the number of African countries in the United Nations, the height of the tallest mountain in Europe and so on. Instead of being asked for a single estimate, however, you were asked to give a 90 per cent confidence interval – a range such that you were 90 per cent confident the true number falls within it. When psychologists run experiments like this, they find that most people give ranges that are too small: the true number falls within their intervals far less than 90 per cent of the time. That is, most people are too sure of their own abilities.

People Give Too Much Weight to a Small Number of Vivid Observations

Imagine that you are thinking about buying a new smart phone by company X. To learn about its reliability, you read *Consumer Reports*, which has surveyed 1,000 owners of the particular smart phone that you are looking at. Then you run into a friend who owns such a phone and she tells you that she is really unhappy with the phone. How do you treat your friend's observation? If you think rationally, you will realize that she has only increased your sample size from 1,000 to 1,001, which does not provide much new information. In addition, a process called the *reticular activation system* (RAS) works to bring your attention to instances of this smart phone – you will suddenly start to notice more of them. The RAS is an automatic mechanism in the brain that brings relevant information to our attention. Both these effects – your friend's story is so vivid and you have noticed more of these smart phones around – mean that you may be tempted to attach a disproportionate weight to them in decision making.

People are Reluctant to Change Their Minds

People tend to interpret evidence to confirm beliefs they already hold. In one study, subjects were asked to read and evaluate a research report on whether capital punishment deters crime. After reading the report, those who initially favoured the death penalty said they were surer in their view, and those who initially opposed the death penalty also said they were surer in their view. The two groups interpreted the same evidence in exactly opposite ways.

People Have a Natural Tendency to Look for Examples Which Confirm Their Existing View or Hypothesis

Nassim Nicholas Taleb, author of the book *The Black Swan*, calls this 'naïve empiricism'. People identify, select or observe past instances and quote them as evidence for a viewpoint or hypothesis. For example, every extreme weather event that is reported is selected as evidence of climate change, or a rise in the price of petrol of 10 per cent is symptomatic of a broader increase in prices of all goods.

People Use Rules of Thumb – Heuristics

The standard economic model implies that to act rationally buyers will consider all available information in making purchasing decisions and weigh up this information to arrive at a decision which maximizes utility subject to the budget constraint. In reality it is likely that many consumers will:

- a. not have access to sufficient information to be able to make a fully rational choice; and
- b. even if they did they would not be able to process this information fully partly due to a lack of mental facility (not everyone can do arithmetic quickly in their head nor make statistical calculations on which to base their choices). Instead when making decisions many people will use short cuts that help simplify the decision-making process. These short cuts are referred to as **heuristics** or *rules of thumb*. Some of these heuristics can be deep-seated and firms can take advantage of them to influence consumer behaviour.

heuristics short cuts or rules of thumb that people use in decision making

There are a number of different types of heuristics.

Anchoring This refers to the tendency for people to start with something they are familiar with or know and make decisions or adjustments based on this anchor. For example, a consumer may base the price they expect to pay for a restaurant meal on the last two prices they paid when eating out. If the price at the next restaurant is higher than this anchor price it may be that the consumer thinks the restaurant is 'expensive' or 'not good value for money' and may choose not to go again, whereas if the price they pay is lower than the anchor price they might see the restaurant as being good value for money and choose to return again. Often these anchors are biased and so the adjustment or decision is flawed in some way.

The Availability Heuristic Cases where decisions are made based on an assessment of the risks of the likelihood of something happening are referred to as availability heuristics. If examples readily come to mind as a result of excessive media coverage, for example, decisions may be taken with a skewed assessment of the risks. If a consumer brings to mind the idea that the last couple of winters have been particularly bad then they might be more likely to buy equipment to help them combat adverse weather for the next winter. Consumers who use commuter trains are more likely to give negative feedback about the service they have received if their recent experience has been of some delays or cancellations even if the overall level of punctuality of the train operator has been very high.

The Representativeness Heuristic In this instance people tend to make judgements by comparing how representative something is to an image or stereotype that they hold. For example, people may be more prepared to pay money to buy a lottery ticket if a close friend has just won a reasonable amount of money on the lottery or make an association that if Bose headphones, for example, are good quality then its home theatre systems are also going to be good quality.

Persuasion Heuristics These are linked to various attributes that a consumer attaches to a product or a brand. For example, it has been shown that size does matter to consumers and so marketers can exploit this by making more exaggerated claims in adverts or using facts and figures to make the product more compelling in the mind of the consumer. The more that the marketers can highlight the positive attributes of their product (and the negative ones of their rivals) the more likely consumers are to make choices in favour of their product. In addition, consumers are also persuaded by people they like and respect. This may be utilized by firms through the people they use in adverts and celebrity endorsements but may also be important in terms of the people a firm employs to represent them in a sales or marketing capacity. It may also be relevant in cases where friends or colleagues talk about products and is one of the reasons why firms are keen to build a better understanding of how social media like Facebook and Twitter can be exploited.

Persuasion heuristics can also manifest themselves in the ‘bandwagon’ effect – if a large number of people go and see a movie and rave about it then there is even more incentive for others to go and see it as well. Firms may look to try and create a bandwagon effect to utilize this persuasion heuristic in their marketing.

Simulation Heuristics These occur where people use mental processes to establish the likely outcome of something. The easier it is to simulate or visualize that outcome the more likely the individual is to make decisions based on it. For example, if it is easy to imagine a product which makes you look good then you are more likely to buy it. Pharmaceutical firms know that consumers are more likely to buy and take medicines that deal with known and experienced symptoms (things like headaches, strained muscles, sore throats and runny noses) which are easy to visualize and imagine than taking regular medicines for something like high cholesterol because it is hard to build a mental process for the effects of high cholesterol.

Expected Utility Theory and Framing Effects

In our analysis of the SEM we noted that indifference curves implied that consumers can rank preferences from best to worst (or vice versa). This is referred to as **expected utility theory**.

expected utility theory the idea that preferences can and will be ranked by buyers

Expected utility theory is important because consumers have to make decisions based on ranking preferences on a regular basis. Imagine you are faced with buying your first car. You have limited income and so have to go to second-hand dealers. You are wary of the potential problems inherent in buying a second-hand car – you know that if the car develops a mechanical fault it often costs more to repair than to replace, so you go to the dealers with this in mind. You find two cars that you like and are faced with the following choice:

Choice (a)

- Car X is priced at €500 and its age suggests that it has a 50 per cent chance of breaking down in the first year.
- Car Y is priced at €900 but only has a 20 per cent chance of breaking down in the first year.

Which do you choose?

Expected utility theory says that consumers can rank the preference between these two options. We assumed that if the car breaks down it is more expensive to repair than to replace so our calculations are based on replacing the car rather than repairing it (we are talking rational human beings after all here). Look at the following choice:

The expected replacement cost of the first car would be the price we paid (€500) × the probability of it breaking down (50 per cent) which works out as $€500 \times 0.50 = €250$. The expected replacement cost of the second car is $0.2 \times 900 = €180$. The rational choice, therefore, would be to purchase the more expensive car. The problem is that the way in which such choices are presented can affect our judgements and the rational decision is violated.

Whilst expected utility theory has been a part of the SEM, research has suggested that there are also flaws in this model. Research into this area is extensive and persuasive. Essentially, choices can be affected by the way in which they are framed. Consider the choice below:

Choice (b)

- Car T has an 80 per cent chance of losing €720 in a year's time.
- Car Z has a 50 per cent chance of losing €250 after a year.

Which would you now choose?

Choice option (b) is exactly the same as choice option (a) but just presented differently – it has been framed differently (the information relating to Car T is the same to that in Car Y in choice (a) and the information relating to Car Z is the same to that in Car X in choice (a)). The risks now appear different.

Firms are careful to frame the way they present products and information to consumers to try to influence purchasing decisions and exploit these differences in perception. For example, firms selling insurance know that people make judgements about the extent to which they are exposed to risk in deciding whether to take out insurance and how much cover they need. Adverts and marketing, therefore, may be framed to give the impression to consumers that they face increased risk.

IN THE NEWS

Neuroscience and Economics

Whilst economists can defend the use of models such as that used in this chapter, they are also aware that new information can also shed light on consumer decision making. The development of magnetic resonance imaging (MRI) is one example of how economists are able to gain further insights into human behaviour and links with other disciplines, notably psychology, is further extending the frontiers of our understanding.

The Brain and Consumer Choice: Neuroeconomics

Two pieces of research show how the work of psychologists and neuroscientists can help to increase our understanding of the processes involved in decision making in consumption. In one piece of research, investigators looked at the role played

by different areas of the brain in making decisions about investing in financial assets. Researchers also looked at the role of the brain in making purchasing decisions related to preference and price in purchasing chocolate. In both cases, the research was investigating how humans balance out anticipation of the gains and losses that we might get from purchases and whether

there was any part of the brain that was associated with being stimulated when faced with the likely gains and losses of a prospective purchase.

The principle behind this investigation was based on the anticipatory effects that occur when faced with a purchasing decision. It is not difficult to conceptualize on a personal level. Think about when you

(Continued)

go into a shop or a store and see something you really like. It might be an expensive purchase and you might know that you should not really be thinking of buying it but the item has really caught your attention and you are debating whether to buy it or not.

On the one hand, you know that the price that is being asked represents a loss – the money you will have to give up is an opportunity cost and means you will have to sacrifice other things that it could also buy you. Equally, you also know that owning this item will provide you with some form of gain – utility. When making the purchasing decision, therefore, you are balancing out these competing forces. The brain plays a role in this; there seem to be two main parts of the brain that are involved in such decisions – the *nucleus accumbens* and the *anterior insular*. The *nucleus accumbens* emits two neuro transmitters – dopamine and serotonin. The former is associated with desire and the latter with inhibition. The *anterior insula* is a part of the brain that has some association with emotional experience and conscious feelings. This will include feelings of pain, happiness, disgust, fear and anger.

The research was relatively complex but the main results from the investigations are given below:

- When making decisions on financial instruments, investors tend not to act rationally. These can be called ‘risk-seeking mistakes’ and ‘risk-aversion mistakes’.
- Activity in the *nucleus accumbens* has an association with risk-seeking mistakes and risky choices.
- Activity in the *anterior insula* is associated with risk-aversion mistakes and riskless choices.
- Distinct neural circuits associated with anticipatory affect lead to different types of choices.

- *Nucleus accumbens* activation represents gain prediction.
- *Anterior insula* activation represents loss prediction.
- Activation of these brain regions can predict decisions to purchase.

In this chapter we looked at a model of consumer behaviour based on a model of rational decision making – maximizing utility with the constraint of a limited income. However, this might not be the case in reality. The expectations that we might have about a purchase might trigger behavioural or affective (to do with emotions) responses which might influence what we eventually end up buying. The findings might be related to a wide range of decision making, including how we choose to buy insurance and assurance, gambling at casinos, at racing and so on, and how we pay for goods and services.

There is a suggestion, for example, that the perceived risk associated with buying goods and services through credit cards is different to that of cash. It has been suggested that we are ‘anaesthetized’ against the effects of paying. As a result there may be a tendency to overspend when purchasing with credit cards compared to cash. There may also be similar effects going on when people purchase online rather than going to a shop or a store. By understanding how the brain works in this respect, there may be possibilities of building incentives or disincentives to purchase to take account of the way these parts of the brain work.

In making purchasing decisions, we know that there are a number of ‘laws’ of economics that we might quote. Such ‘laws’ allow us to be able to make predictions – the very basis of having theories. However, if these laws are based on inaccurate or incomplete knowledge of how humans actually carry out these

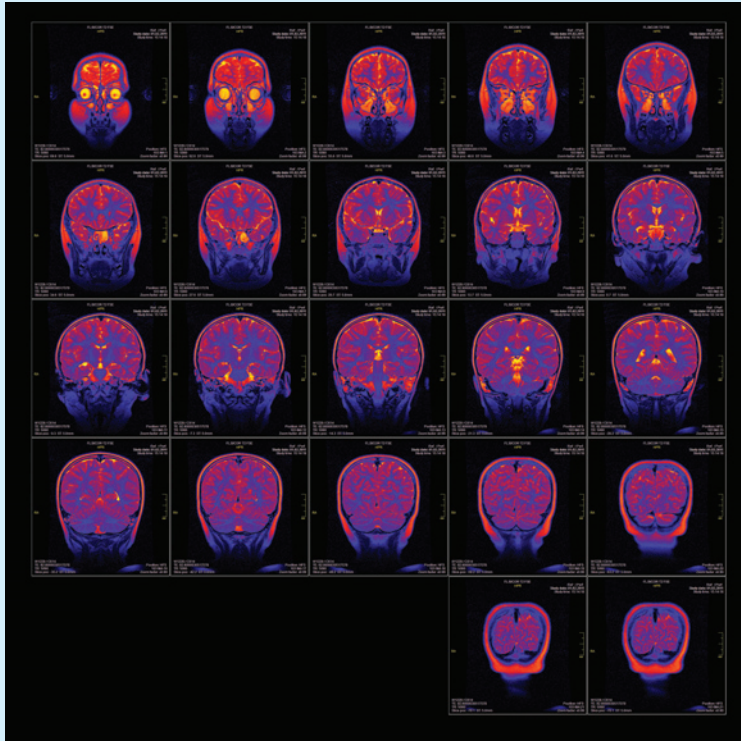
actions then the model will not be able to be used as a predictive tool.

We assume, for example, that humans are attracted to preferred products (how that preference is generated is another story). We also assume that consumers prefer lower prices than higher prices and that if a price is deemed ‘excessive’ then we will avoid purchasing the item. Prices represent a potential gain and a potential loss. However, we might view similar losses and gains in different ways. In some cases we might view a loss as being more important than a gain even if they were of equal magnitude. That contradicts the rational approach that characterizes much of economic theory. If our decision making is being affected by the activation of these distinct neural circuits then the extent to which they are activated might influence our choice. For example, the research cited examples where men were shown pictures of sports cars and other types of car deemed to be less desirable. Viewing sports cars produced a greater degree of brain activity in the mid-brain. Similar results have been obtained in cases where both men and women are shown preferred rather than less preferred drinks, brands of beer and coffee.

If this is the case, then it might suggest that advertising over a long period of time might have significant effects in influencing our choices if it has effects on our perceptions of preferred or non-preferred purchases. What is happening is that the perception of loss and gain is being altered in relation to the price that is being charged.

The second piece of research involves the effect on human behaviour of different emotional states and the effect that hormones play in influencing the brain and human behaviour. In a paper presented to a British Psychological Society (BPS) meeting

The brain plays a crucial role in decision making; new technology such as MRI scans can highlight activity in different parts of the brain when decision making is taking place increasing our understanding of human behaviour.



in April 2009, Professor Karen Pine of the University of Hertfordshire outlined her research findings from a study looking at the link between the menstrual cycle in women and purchasing decisions. Professor Pine's study involved 443 women aged 18 to 50. It appeared to show a link between the stage of the menstrual cycle and purchasing decisions. Of the women in the sample, 153 were in the later stages of their menstrual cycle, known as the luteal phase. Of this group, over 60 per cent said that they had indulged in overspending and had bought items on impulse. The spending, in a small number of cases, was way over normal budgets – some women saying that they had overspent by as much as £250. In many cases, the women said that their purchasing decisions at this time were accompanied by feelings of remorse afterwards.

Professor Pine commented, 'The spending behaviour tends to be a reaction to intense emotions. They are feeling stressed or depressed and are more likely to go shopping to cheer themselves up and using it to regulate their emotions.' The study also found that those women with severe pre-menstrual tension (PMT) displayed more extreme examples of this behaviour.

Part of the explanation for this behaviour is that hormonal changes in women at certain times of the menstrual cycle are associated with changes to the part of the brain that is linked to inhibitions and emotions. It has also been suggested that purchasing decisions may be linked to a desire of women to make themselves look more attractive. There has been other research which suggested that part of female behaviour is driven by the need to demonstrate their fertility – a

throwback to very early days when survival of the species depended in part on the ability to reproduce successfully. Around 14 days before the start of ovulation is a time which sees women increase their spending on items that enhance their attractiveness. These items include make-up, high-heeled shoes and jewellery. Professor Pine has said that her findings are supported by other research which shows a so-called 'ornamental effect' linked to stages in the menstrual cycle.

Questions

- 1 To what extent should we assume that consumers can accurately estimate the potential gains and losses of a purchase decision (i.e. the utility)? How does your answer to this question affect your view of the SEM?**
- 2 Think of a purchase you have made in recent months after which you experienced guilt. Explain why you felt guilty and whether such feelings can be accounted for in the SEM.**
- 3 If research suggests that consumer behaviour is far more complex than the SEM suggests, does that mean the SEM is of no value and study of it ought to be discarded from undergraduate economics courses? Explain.**
- 4 Explain the role that emotion might play in consumer decision making. (You should think about the range of emotions that have been alluded to in the article.)**
- 5 Some economists have suggested that human brains have not evolved to cope with the rapid changes in our lifestyles – that our brains are still hard-wired to cope with the need to survive, characteristic of early humans. Using the second piece of research outlined in the article, explain how such a theory might affect our understanding of consumer behaviour.**

SUMMARY

- The analysis of consumer choice looks at how consumers make decisions. There are a number of assumptions underpinning the model which include that people behave rationally to maximize their utility from their given resources.
- A consumer's budget constraint shows the possible combinations of different goods he can buy given his income and the prices of the goods. The slope of the budget constraint equals the relative price of the goods.
- The consumer's indifference curves represent his preferences. An indifference curve shows the various bundles of goods that make the consumer equally happy. Points on higher indifference curves are preferred to points on lower indifference curves. The slope of an indifference curve at any point is the consumer's marginal rate of substitution – the rate at which the consumer is willing to trade one good for the other.
- The consumer optimizes by choosing the point on his budget constraint that lies on the highest indifference curve. At this point, the slope of the indifference curve (the marginal rate of substitution between the goods) equals the slope of the budget constraint (the relative price of the goods).
- When the price of a good falls, the impact on the consumer's choices can be broken down into an income effect and a substitution effect. The income effect is the change in consumption that arises because a lower price makes the consumer better off. The substitution effect is the change in consumption that arises because a price change encourages greater consumption of the good that has become relatively cheaper. The income effect is reflected in the movement from a lower to a higher indifference curve, whereas the substitution effect is reflected by a movement along an indifference curve to a point with a different slope.
- The theory of consumer choice can be applied in many situations. It can explain why demand curves can potentially slope upward, why higher wages could either increase or decrease the quantity of labour supplied, and why higher interest rates could either increase or decrease saving.

QUESTIONS FOR REVIEW

- 1 What are the main assumptions of the standard economic model?
- 2 A consumer has income of €3,000. Wine is priced at €3 a glass and cheese is priced at €6 a kilo. Draw the consumer's budget constraint. What is the slope of this budget constraint?
- 3 Draw a consumer's indifference curves for wine and cheese. Describe and explain four properties of these indifference curves.
- 4 Pick a point on an indifference curve for wine and cheese and show the marginal rate of substitution. What does the marginal rate of substitution tell us?
- 5 Show a consumer's budget constraint and indifference curves for wine and cheese. Show the optimal consumption choice. If the price of wine is €3 a glass and the price of cheese is €6 a kilo, what is the marginal rate of substitution at this optimum?
- 6 A person who consumes wine and cheese gets a rise, so his income increases from €3,000 to €4,000. Use diagrams to show what happens if both wine and cheese are normal goods. Now show what happens if cheese is an inferior good.
- 7 The price of cheese rises from €6 to €10 a kilo, while the price of wine remains at €3 a glass. For a consumer with a constant income of €3,000, show what happens to consumption of wine and cheese. Decompose the change into income and substitution effects.
- 8 Can an increase in the price of cheese possibly induce a consumer to buy more cheese? Explain.
- 9 Explain why the assumptions of the standard economic model might not hold.
- 10 What are heuristics and how might they affect consumer decision making?

PROBLEMS AND APPLICATIONS

- 1 Jacqueline divides her income between coffee and croissants (both of which are normal goods). An early frost in Brazil causes a large increase in the price of coffee in France.
 - a. Show how this early frost might affect Jacqueline's budget constraint.
 - b. Show how this early frost might affect Jacqueline's optimal consumption bundle, assuming that the substitution effect outweighs the income effect for croissants.
 - c. Show how this early frost might affect Jacqueline's optimal consumption bundle, assuming that the

income effect outweighs the substitution effect for croissants.

2 Compare the following two pairs of goods:

- a. Coke and Pepsi
- b. skis and ski bindings.

In which case do you expect the indifference curves to be fairly straight, and in which case do you expect the indifference curves to be very bowed? In which case will the consumer respond more to a change in the relative price of the two goods?

3 Surette buys only orange juice and yoghurt.

- a. In 2014, Surette earns €20,000, orange juice is priced at €2 a carton and yoghurt is priced at €4 a tub. Draw Surette's budget constraint.
- b. Now suppose that all prices increase by 10 per cent in 2015 and that Surette's salary increases by 10 per cent as well. Draw Surette's new budget constraint. How would Surette's optimal combination of orange juice and yoghurt in 2015 compare to her optimal combination in 2014?

4 Economist George Stigler once wrote that, according to consumer theory, 'if consumers do not buy less of a commodity when their incomes rise, they will surely buy less when the price of the commodity rises'. Explain this statement using the concepts of income and substitution effects.

5 A consumer has an income of €30,000 a year and divides this income between buying food and spending on leisure. The average price of a unit of food is €15 and the average price of a unit of leisure is €10. Draw the consumer's budget constraint and draw an indifference curve to show the consumer's optimum.

Assume that the price of food rises in three stages over the year in €10 increments but that the price of leisure

stays the same. Draw the consumer's new budget constraints, identify the new optimum and show the price-consumption path. Use the price-consumption path to derive the demand curve for food.

6 Indifference curves are concave to the origin (i.e. they bow outwards). Using your knowledge of the properties of indifference curves, explain why indifference curves cannot be convex to the origin (i.e. bow inwards).

7 Using an example, explain why the consumer optimum occurs where the ratio of the prices of two goods is equal to the marginal rate of substitution.

8 Sketch a diagram to show the effect on demand of a change in income on a good which has an income elastic demand.

9 Choose three products you purchased recently. Think about the reasons that you made the particular purchase decision in each case. To what extent do you think you made the purchase decision in each case in line with the assumptions of the SEM? If you deviated from the SEM, think about why you did so.

10 Look at the following two statements:

- a. Which would you prefer – a 50 per cent chance of winning €150 or a 50 per cent chance of winning €100?
- b. Would you prefer a decision that guarantees a €100 loss or would you rather take a gamble where the chance of winning €50 was rated at 50 per cent but the chance of losing €200 rated also at 50 per cent? What would your choice be in a? What would your choice be in b?

What is the difference between these two sets of statements and how do they illustrate the concept of framing?

6

BACKGROUND TO SUPPLY: FIRMS IN COMPETITIVE MARKETS

In this chapter we will examine the behaviour of competitive firms which will help explain the decisions that lie behind the supply curve. You may recall that a market is competitive if each buyer and seller is small compared to the size of the market and, therefore, has little ability to influence market prices. We will find that a market supply curve is tightly linked to firms' costs of production. But among a firm's various costs – fixed, variable, average and marginal – which ones are most relevant for its decision about the quantity to supply at any given price? We will see that all these measures of cost play important and interrelated roles.

THE COSTS OF PRODUCTION

All firms, from Air France to your local baker's shop, incur costs as they make the goods and services that they sell. A firm's costs are a key determinant of its production and pricing decisions. In this chapter, we define some of the variables that economists use to measure a firm's costs, and we consider the relationships among them. It provides a crucial foundation for the fascinating topics that follow.

What Are Costs?

We will look at the costs of a firm using an example – Paolo's Pizza Factory. Paolo, the owner of the firm, buys flour, tomatoes, mozzarella cheese, salami and other pizza ingredients. He also buys the mixers and ovens, and hires workers to run this equipment. He then sells the resulting pizzas to consumers. By examining some of the issues that Paolo faces in his business, we can learn some lessons about costs that apply to all firms in the economy.

Costs as Opportunity Costs

When measuring costs at Paolo's Pizza Factory, or any other firm, it is important to keep in mind one of the *Ten Principles of Economics* from Chapter 1: the cost of something is what you give up to get it. Recall that the *opportunity cost* of an item refers to all those things that must be forgone to acquire that item. When economists speak of a firm's cost of production, they include the opportunity costs of making its output of goods and services.

A firm's opportunity costs of production are sometimes obvious and sometimes less so. When Paolo pays €1,000 for flour, Paolo can no longer use that €1,000 to buy something else; he has to sacrifice what else that €1,000 could have purchased. Similarly, when Paolo hires workers to make the pizzas, the wages he pays are part of the firm's costs. Because these costs require the firm to pay out some money, they are called **explicit costs**. By contrast, some of a firm's opportunity costs, called **implicit costs**, do not require a cash outlay. Imagine that Paolo is skilled with computers and could earn €100 per hour working as a programmer. For every hour that Paolo works at his pizza factory, he gives up €100 in income, and this forgone income is also classed as part of his costs by an economist.

explicit costs input costs that require an outlay of money by the firm
implicit costs input costs that do not require an outlay of money by the firm

This distinction between explicit and implicit costs highlights an important difference between how economists and accountants analyse a business. Economists are interested in studying how firms make production and pricing decisions. Because these decisions are based on both explicit and implicit costs, economists include both when measuring a firm's costs. By contrast, accountants have the job of keeping track of the money that flows into and out of firms. As a result, they measure the explicit costs but often ignore the implicit costs.

The difference between economists and accountants is easy to see in the case of Paolo's Pizza Factory. When Paolo gives up the opportunity to earn money as a computer programmer, his accountant will not count this as a cost of his pizza business. Because no money flows out of the business to pay for this cost, it never shows up on the accountant's financial statements. An economist, however, will count the foregone income as a cost because it will affect the decisions that Paolo makes in his pizza business. This is an important part of thinking like an economist. In Paolo's case, if the wage as a computer programmer rose from €100 to €500 per hour, the opportunity cost of running his pizza business might now change his decision making. The opportunity cost of running the business in terms of what Paolo is sacrificing in foregone income has risen. Paolo might decide he could earn more by closing the business and switching to computer programming.

The Cost of Capital as an Opportunity Cost

An important implicit cost of almost every business is the opportunity cost of the financial capital that has been invested in the business. The difference in the way economists and accountants treat costs is especially true in their treatment of the cost of capital. Suppose, for instance, that Paolo used €300,000 of his savings to buy his pizza factory from the previous owner. If Paolo had instead left this money deposited in a savings account that pays an interest rate of 5 per cent, he would have earned €15,000 per year (assuming simple interest). To own his pizza factory, therefore, Paolo has given up €15,000 a year in interest income. This forgone €15,000 is an implicit opportunity cost of Paolo's business. An economist views the €15,000 in interest income that Paolo gives up every year as a cost of his business, even though it is an implicit cost. Paolo's accountant, however, will not show this €15,000 as a cost because no money flows out of the business to pay for it.

To explore further the difference between economists and accountants, let's change the example slightly. Suppose now that Paolo did not have the entire €300,000 to buy the factory but, instead, used €100,000 of his own savings and borrowed €200,000 from a bank at an interest rate of 5 per cent. Paolo's accountant, who only measures explicit costs, will now count the €10,000 interest paid on the bank loan every year as a cost because this amount of money now flows out of the firm. By contrast, according to an economist, the opportunity cost of owning the business is still €15,000. The opportunity cost equals the interest on the bank loan (an explicit cost of €10,000) plus the forgone interest on savings (an implicit cost of €5,000).

SELF TEST Richard Collishaw is a farmer who is also a skilled metal worker. He makes unique garden sculptures that could earn him €40 an hour. One day, he spends 10 hours planting €500 worth of seeds on his farm. What opportunity cost has he incurred? What cost would his accountant measure? If these seeds will yield €1,000 worth of crops, does Richard earn an accounting profit? Does he earn an economic profit? Would you advise Richard to continue as a farmer or switch to metal working?

PRODUCTION AND COSTS

Firms incur costs when they buy inputs to produce the goods and services that they plan to sell. In the analysis that follows, we make an important simplifying assumption: we assume that the size of Paolo's factory is fixed so that Paolo can vary the quantity of pizzas produced only by changing the number of

workers. This assumption is realistic in **the short run**, but not in **the long run**. That is, Paolo cannot build a larger factory overnight, but he can do so within a year or so. This initial analysis, therefore, should be viewed as describing the production decisions that Paolo faces in the short run.

the short run the period of time in which some factors of production cannot be changed
the long run the period of time in which all factors of production can be altered

The Production Function

Table 6.1 shows how the quantity of pizzas Paolo's factory produces per hour depends on the number of workers. As you see in the first two columns, if there are no workers in the factory Paolo produces no pizzas. When there is 1 worker he produces 50 pizzas. When there are 2 workers he produces 90 pizzas, and so on. Panel (a) of Figure 6.1 presents a graph of these two columns of numbers. The number of workers is on the horizontal axis, and the number of pizzas produced is on the vertical axis. This relationship between the quantity of inputs (workers) and quantity of output (pizzas) is called the **production function**.

production function the relationship between the quantity of inputs used to make a good and the quantity of output of that good

TABLE 6.1 A Production Function and Total Cost: Paolo's Pizza Factory

| Number of workers | Output (quantity of pizzas produced per hour) or total product | Marginal product of labour | Cost of factory (€) | Cost of workers (€) | Total cost of inputs (cost of factory + cost of workers) (€) |
|-------------------|--|----------------------------|---------------------|---------------------|--|
| 0 | 0 | 0 | 30 | 0 | 30 |
| 1 | 50 | 50 | 30 | 10 | 40 |
| 2 | 90 | 40 | 30 | 20 | 50 |
| 3 | 120 | 30 | 30 | 30 | 60 |
| 4 | 140 | 20 | 30 | 40 | 70 |
| 5 | 150 | 10 | 30 | 50 | 80 |

The production function can be represented as a mathematical function where output (Q) is dependent on the factor inputs capital (K) and labour (L):

$$Q = f(K, L)$$

We have come across marginal concepts in Chapter 3 and this concept is applicable to understanding the decision a firm makes about how many workers to hire and how much output to produce. The third column in Table 6.1 gives the marginal product of a worker. The **marginal product** of any factor input (MP_F) in the production process is the increase in the quantity of output obtained from one additional unit of that factor input represented as:

$$MP_F = \frac{\text{change in total product}}{\text{change in quantity of the factor}}$$

marginal product the increase in output that arises from an additional unit of input

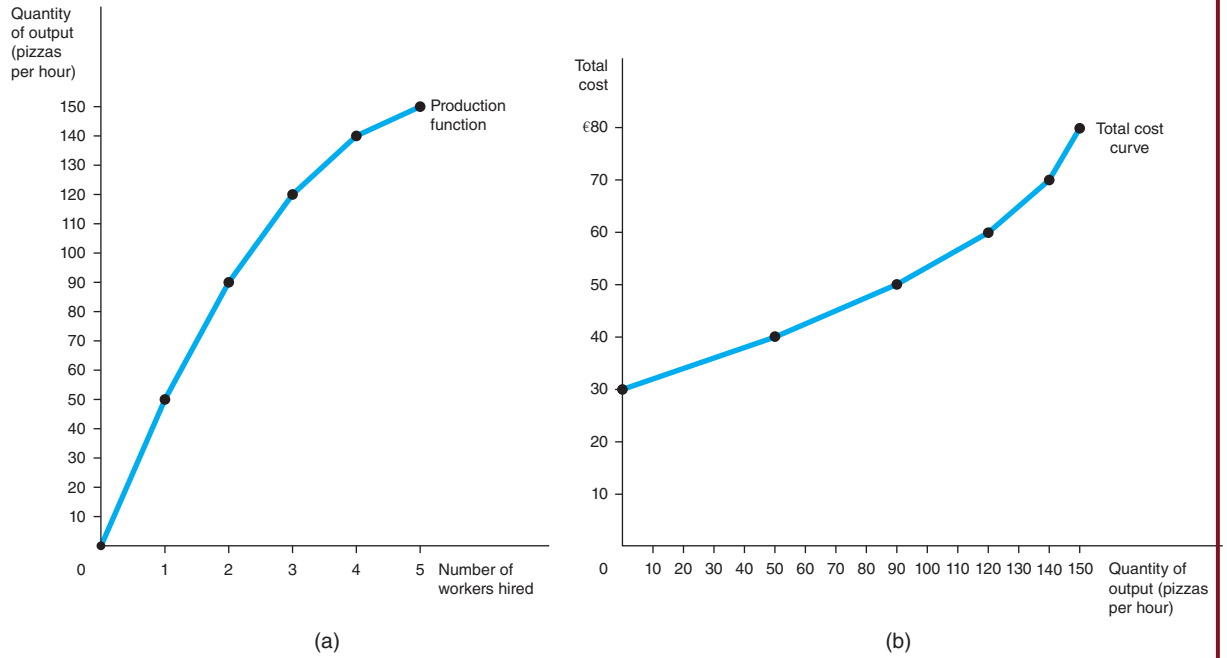
The marginal product of labour, therefore, would be represented as:

$$MP_L = \frac{\Delta Q}{\Delta L}$$

FIGURE 6.1

Paolo's Production Function

The production function in panel (a) shows the relationship between the number of workers hired and the quantity of output produced. Here the number of workers hired (on the horizontal axis) is from the first column in Table 6.1, and the quantity of output produced (on the vertical axis) is from the second column. The production function (total product) gets flatter as the number of workers increases, which reflects diminishing marginal product. The total cost curve in panel (b) shows the relationship between the quantity of output produced and total cost of production. Here the quantity of output produced (on the horizontal axis) is from the second column in Table 6.1, and the total cost (on the vertical axis) is from the sixth column. The total cost curve gets steeper as the quantity of output increases because of diminishing marginal product.



If we were using calculus the marginal product for an infinitesimal small change in labour input would be the partial derivative:

$$MP_L = \frac{\partial Q}{\partial L}$$

We use partial derivatives in this case because the production function has more than one independent variable (capital and labour). We are looking at the rate of change of Q when either K or L varies but with the other variables held constant. To find the marginal product of labour, therefore, we are differentiating Q with respect to L whilst K is held constant.

When the number of workers goes from 1 to 2, total pizza production increases from 50 to 90, so the marginal product of the second worker is 40 pizzas. When the number of workers goes from 2 to 3, total product increases from 90 to 120, so the marginal product of the third worker is 30 pizzas.

Notice that as the number of workers increases, the marginal product (MP) declines. The second worker has a MP of 40 pizzas, the third worker has a MP of 30 pizzas and the fourth worker has a MP of 20 pizzas. This property of the production function is called **diminishing marginal product**. Why does this happen? At first, when only a few workers are hired, they have easy access to Paolo's kitchen equipment. As the number of workers increases, additional workers have to share equipment and work in more crowded conditions, and if the number of workers continued to be increased the workers would start to get in each

diminishing marginal product the property whereby the marginal product of an input declines as the quantity of the input increases

other's way and efficiency would be significantly impaired. Hence, as more and more workers are hired, each additional worker contributes less to the production of pizzas.

Diminishing marginal product is also apparent in panel (a) of Figure 6.1. The production function's slope ('rise over run') tells us the change in Paolo's output of pizzas ('rise') for each additional input of labour ('run'). That is, the slope of the production function measures the marginal product of a worker. As the number of workers increases, the marginal product declines, and the production function becomes flatter. In this example the production function is curvilinear and as a result the use of calculus to find marginal product would give a more accurate answer than would simply finding the average of the slope between two points.

For example, suppose that the production function for Paolo's factory looked like this:

$$Q = 20K^{0.5}L^{0.5}$$

If $K = 16$ and $L = 25$, then the marginal product of labour would be:

$$MP_L = \frac{\partial Q}{\partial L}$$

$$\frac{\partial Q}{\partial L} = 10K^{0.5}L^{-0.5} = \frac{10\sqrt{K}}{\sqrt{L}} = \frac{40}{5} = 8$$

From the Production Function to the Total Cost Curve

The last three columns of Table 6.1 are reproduced as a graph in panel (b) of Figure 6.1 to show Paolo's cost of producing pizzas. In this example, the cost of operating the factory is €30 per hour and the cost of a worker is €10 per hour. If Paolo hires 1 worker, his total cost is €40. If he hires 2 workers, his total cost is €50 and so on. With this information, the table now shows how the number of workers Paolo hires is related to the quantity of pizzas he produces and to his total cost of production.

An important relationship in Table 6.1 is between quantity produced (in the second column) and total costs (in the sixth column). Panel (b) of Figure 6.1 graphs these two columns of data with the quantity produced on the horizontal axis and total cost on the vertical axis. This graph is called the *total cost curve*.

Now compare the total cost curve in panel (b) of Figure 6.1 with the production function in panel (a). The total cost of producing the quantity Q is the sum of all production factors where P_L is the price of labour per hour and P_K is the price of hiring capital. $C(Q) = P_L \times L(Q) + P_K \times K(Q)$. Here $L(Q)$ and $K(Q)$ are the labour hours and the amount of capital employed to produce Q units of output. These two curves are opposite sides of the same coin. The total cost curve gets steeper as the amount produced rises, whereas the production function gets flatter as production rises. These changes in slope occur for the same reason. High production of pizzas means that Paolo's kitchen is crowded with many workers. Because the kitchen is crowded, each additional worker adds less to production, reflecting diminishing marginal product. Therefore, the production function is relatively flat. But now turn this logic around: when the kitchen is crowded, producing an additional pizza requires a lot of additional labour and is thus very costly. Therefore, when the quantity produced is large, the total cost curve is relatively steep.

SELF TEST If a farmer plants no seeds on his farm, he gets no harvest. If he plants 1 bag of seeds he gets 5 tonnes of wheat. If he plants 2 bags he gets 7 tonnes. If he plants 3 bags he gets 8 tonnes. A bag of seeds is priced at €100, and seeds are his only cost. Use these data to graph the farmer's production function and total cost curve. Explain their shapes.

THE VARIOUS MEASURES OF COST

Our analysis of Paolo's Pizza Factory demonstrate how a firm's total cost reflects its production function. From data on a firm's total cost we can derive several related measures of cost to help in analysing production and pricing decisions. Consider the example in Table 6.2. This table presents cost data on Paolo's neighbour: Luciano's Lemonade Stand.

The first column of the table shows the number of glasses of lemonade that Luciano might produce, ranging from 0 to 10 glasses per hour. The second column shows Luciano's total cost of producing glasses of lemonade. Figure 6.2 plots Luciano's total cost curve. The quantity of lemonade (from the first column)

TABLE 6.2

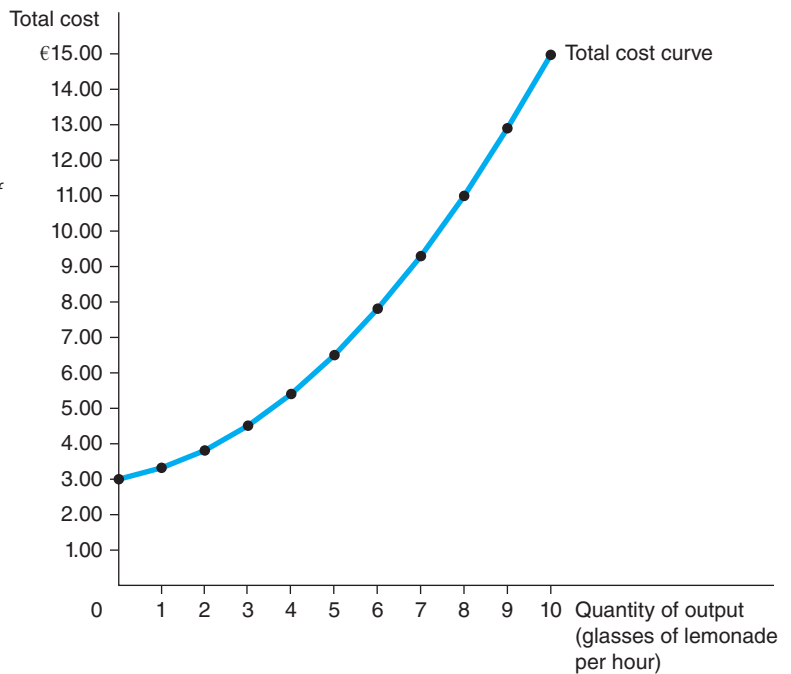
The Various Measures of Cost: Luciano's Lemonade Stand

| Quantity of lemonade glasses (per hour) | Total cost (€) | Fixed cost (€) | Variable cost (€) | Average fixed cost (€) | Average variable cost (€) | Average total cost (€) | Marginal cost (€) |
|---|----------------|----------------|-------------------|------------------------|---------------------------|------------------------|-------------------|
| 0 | 3.00 | 3.00 | 0.00 | – | – | – | 0.30 |
| 1 | 3.30 | 3.00 | 0.30 | 3.00 | 0.30 | 3.30 | 0.50 |
| 2 | 3.80 | 3.00 | 0.80 | 1.50 | 0.40 | 1.90 | 0.70 |
| 3 | 4.50 | 3.00 | 1.50 | 1.00 | 0.50 | 1.50 | 0.90 |
| 4 | 5.40 | 3.00 | 2.40 | 0.75 | 0.60 | 1.35 | 1.10 |
| 5 | 6.50 | 3.00 | 3.50 | 0.60 | 0.70 | 1.30 | 1.30 |
| 6 | 7.80 | 3.00 | 4.80 | 0.50 | 0.80 | 1.30 | 1.50 |
| 7 | 9.30 | 3.00 | 6.30 | 0.43 | 0.90 | 1.33 | 1.70 |
| 8 | 11.00 | 3.00 | 8.00 | 0.38 | 1.00 | 1.38 | 1.90 |
| 9 | 12.90 | 3.00 | 9.90 | 0.33 | 1.10 | 1.43 | 2.10 |
| 10 | 15.00 | 3.00 | 12.00 | 0.30 | 1.20 | 1.50 | |

FIGURE 6.2

Luciano's Total Cost Curve

Here the quantity of output produced (on the horizontal axis) is from the first column in Table 6.2, and the total cost (on the vertical axis) is from the second column. As in Figure 6.2, the total cost curve gets steeper as the quantity of output increases because of diminishing marginal product.



is on the horizontal axis, and total cost (from the second column) is on the vertical axis. Luciano's total cost curve has a shape similar to Paolo's. In particular, it becomes steeper as the quantity produced rises, which (as we have discussed) reflects diminishing marginal product.

Fixed and Variable Costs

Luciano's total cost can be divided into two types. Some costs, called **fixed costs**, are not determined by the amount of output produced; they can change but not as a result of changes in the amount produced. They are incurred even if the firm produces nothing at all. Luciano's fixed costs include any rent he pays because this cost is the same regardless of how much lemonade Luciano produces. Similarly, if Luciano needs to

fixed costs costs that are not determined by the quantity of output produced

hire a bar person to serve the drinks, regardless of the quantity of lemonade sold, the bar person's salary is a fixed cost. The third column in Table 6.2 shows Luciano's fixed cost (FC), which in this example is €3.00.

Some of the firm's costs, called **variable costs**, change as the firm alters the quantity of output produced. Luciano's variable costs include the cost of lemons, sugar, paper cups and straws: the more lemonade Luciano makes, the more of these items he needs to buy. Similarly, if Luciano pays his workers overtime to make more lemonade, the wages of these workers are variable costs. The fourth column of the table shows Luciano's variable cost. The variable cost is 0 if he produces nothing, €0.30 if he produces 1 glass of lemonade, €0.80 if he produces 2 glasses and so on.

variable costs costs that are dependent on the quantity of output produced

A firm's total cost is the sum of fixed and variable costs. In Table 6.2 total cost in the second column equals fixed cost in the third column plus variable cost in the fourth column.

$$TC(Q) = VC(Q) + FC$$

Average and Marginal Cost

As the owner of his firm, Luciano has to decide how much lemonade to produce. A key part of this decision is how his costs will vary as he changes the level of production. In making this decision, Luciano might ask the following two questions about the cost of producing lemonade:

- How much does it cost to make the typical glass of lemonade?
- How much does it cost to increase production of lemonade by 1 glass?

Although at first these two questions might seem to have the same answer, they do not.

To find the cost of the typical unit produced, we would divide the firm's total costs by the quantity of output it produces. For example, if the firm produces 2 glasses per hour, its total cost is €3.80, and the cost of the typical glass is €3.80/2, or €1.90. Total cost divided by the quantity of output is called **average total cost** (ATC).

$$ATC = \frac{TC}{Q}$$

average total cost total cost divided by the quantity of output

Because total cost is just the sum of fixed and variable costs, average total cost can be expressed as the sum of average fixed cost and average variable cost. **Average fixed cost** is the fixed costs (FC) divided by the quantity of output:

$$AFC = \frac{FC}{Q}$$

and **average variable cost** is the variable cost divided by the quantity of output:

$$AVC = \frac{VC}{Q}$$

average fixed cost fixed costs divided by the quantity of output

average variable cost variable costs divided by the quantity of output

The last column in Table 6.2 shows the amount that total cost rises when the firm increases production by 1 unit of output. This number is called **marginal cost**. For example, if Luciano increases production from 2 to 3 glasses, total cost rises from €3.80 to €4.50, so the marginal cost of the third glass of lemonade is €4.50 minus €3.80 or €0.70.

$$MC = \frac{\Delta TC}{\Delta Q}$$

Using calculus:

$$MC = \frac{dTC}{dQ}$$

marginal cost the increase in total cost that arises from an extra unit of production

We can derive various average cost measures and marginal cost from the total cost function of the type $TC = f(Q)$.

If the total cost function is given as $TC = 7Q^2 + 5Q + 1,500$ we can see that the fixed costs are 1,500. If $Q = 0$, then TC would be 1,500. The terms $7Q^2 + 5Q$ are the variable costs, therefore.

AC would be $\frac{7Q^2 + 5Q + 1,500}{Q}$, AVC would be $\frac{7Q^2 + 5Q}{Q}$ and AFC would be $\frac{1,500}{Q}$. Marginal cost would be $\frac{dTC}{dQ} = 14Q + 5$.

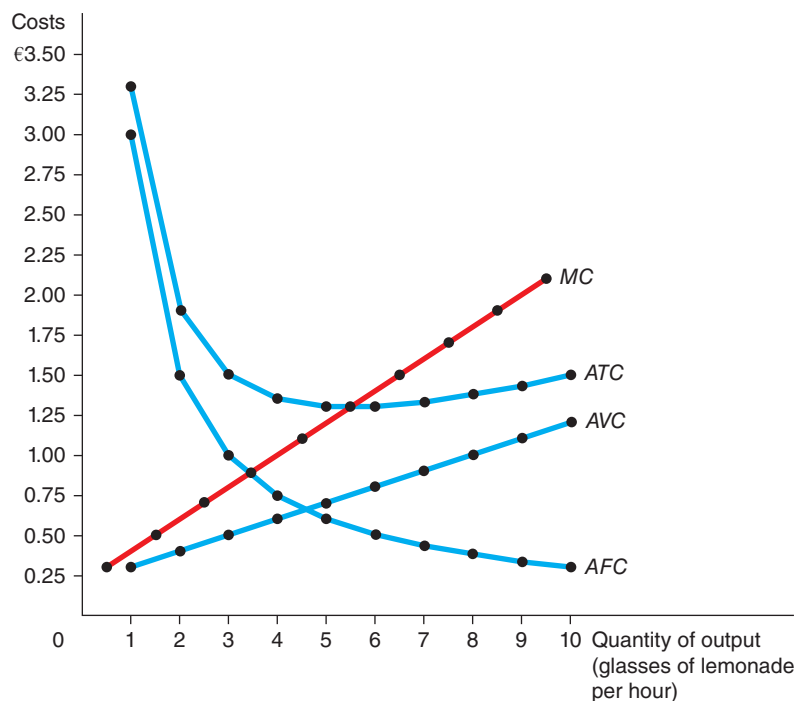
Cost Curves and their Shapes

Just as in previous chapters we found graphs of supply and demand useful when analysing the behaviour of markets, we will find graphs of average and marginal cost useful when analysing the behaviour of firms. Figure 6.3 graphs Luciano's costs using the data from Table 6.2. The horizontal axis measures the quantity the firm produces, and the vertical axis measures marginal and average costs. The graph shows four curves: average total cost (ATC), average fixed cost (AFC), average variable cost (AVC) and marginal cost (MC).

FIGURE 6.3

Luciano's Average Cost and Marginal Cost Curves

This figure shows the average total cost (ATC), average fixed cost (AFC) average variable cost (AVC) and marginal cost (MC) for Luciano's Lemonade Bar. All of these curves are obtained by graphing the data in Table 6.2. These cost curves show three features that are typical of many firms: (1) marginal cost rises with the quantity of output; (2) the average total cost curve is U-shaped; (3) the marginal cost curve crosses the average total cost curve at the minimum of average total cost.



The cost curves shown here for Luciano's Lemonade Bar have some features that are common to the cost curves of many firms in the economy. Let's examine three features in particular: the shape of marginal cost, the shape of average total cost, and the relationship between marginal and average total cost.

Rising Marginal Cost Luciano's marginal cost rises with the quantity of output produced. This reflects the property of diminishing marginal product. When Luciano is producing a small quantity of lemonade he has spare capacity and can easily put these idle resources to use, the marginal product of an extra worker is large, and the marginal cost of an extra glass of lemonade is small. By contrast, when Luciano is producing a large quantity of lemonade his stand is crowded with workers and most of his equipment is fully utilized. Luciano can produce more lemonade by adding workers, but these new workers have to work in crowded conditions and may have to wait to use the equipment. Therefore, when the quantity of lemonade being produced is already high, the marginal product of an extra worker is low, and the marginal cost of an extra glass of lemonade is large.

U-Shaped Average Total Cost Luciano's average total cost curve takes on a U-shape. To understand why this is so, remember that average total cost is the sum of average fixed cost and average variable cost. Average fixed cost always declines as output rises because the fixed cost does not change as output rises and so gets spread over a larger number of units. Average variable cost typically rises as output increases because of diminishing marginal product. Average total cost reflects the shapes of both average fixed cost and average variable cost. As shown in Figure 6.3, at very low levels of output, such as 1 or 2 glasses per hour, average total cost is high because the fixed cost is spread over only a few units. Average total cost then declines as output increases until the firm's output reaches 5 glasses of lemonade per hour, when average total cost falls to €1.30 per glass. When the firm produces more than 6 glasses, average total cost starts rising again because average variable cost rises substantially. If further units of output were produced the average total cost curve would continue to slope upwards giving the typical U-shape referred to.

The bottom of the U-shape occurs at the quantity that minimizes average total cost. This quantity is sometimes called the **efficient scale** of the firm or *minimum efficient scale*. For Luciano, the efficient scale is 5 or 6 glasses of lemonade. If he produces more or less than this amount, his average total cost rises above the minimum of €1.30.

efficient scale the quantity of output that minimizes average total cost

The Relationship between Marginal Cost and Average Total Cost Whenever marginal cost is less than average total cost, average total cost is falling. Whenever marginal cost is greater than average total cost, average total cost is rising. This feature of Luciano's cost curves is not a coincidence from the particular numbers used in the example: it is true for all firms and is a basic mathematical relationship.

To see why, refer to your understanding of averages and consider what happens to average cost as output goes up by one unit. If the cost of the extra unit is (say) €5 and the cost of the next unit produced is €3, what will happen to the average? The answer is that the average will fall. If the cost of the additional unit was €8 then the average will increase. The cost of an extra unit is the marginal cost, so it follows that if marginal cost is less than average cost, average cost will be falling; and if marginal cost is above average cost, average cost will be rising.

This relationship between average total cost and marginal cost has an important corollary: the marginal cost curve crosses the average total cost curve at its minimum. Why? At low levels of output, marginal cost is below average total cost, so average total cost is falling. But after the two curves cross, marginal cost rises above average total cost. For the reason we have just discussed, average total cost must start to rise at this level of output. Hence, at this point of intersection the cost of an additional unit is the same as the average and so the average does not change and the point is the minimum of average total cost.

CASE STUDY Understanding Averages and Marginal Values

When first encountering average and marginal costs, it can be a little confusing, so it is often worth thinking about these values in a different context. Sebastian is an up-and-coming basketball player. In the first 18 games of the season Sebastian has scored a total of 306 points. The average points per game (PPG) at this stage of the season, therefore, is $306/18 = 17$ points per game. The number of points that Sebastian scores in each successive game is his marginal score. If Sebastian scores 15 points in the 19th game his total points score is now 321 and so his average will now fall to $321/19 = 16.89$. If in the next three games he scores 12, 14 and 10 then the marginal points score continues to be below the average and so his average will continue to fall. At the end of the 22nd game he will have scored a total of 357 points and his average will have fallen to 16.22. If in the next four games he scores 20, 18, 22 and 20 points, then each successive game sees his marginal points score being better than his average and so the average will be pulled up. At the end of 26 games, therefore, his total points score will be 437 and his average PPG will have risen to 16.8.

So, if Sebastian's marginal points score in each successive game is less than his average, his average will fall but if his marginal points score in each game is more than his average points score then his average will rise. It's the same principle for firms' costs!



Thinking about average and marginal costs by using a context such as sport helps develop a deeper understanding of the conceptual and mathematical relationship between the two. We have used basketball but you can apply it to almost any sport.

Typical Cost Curves

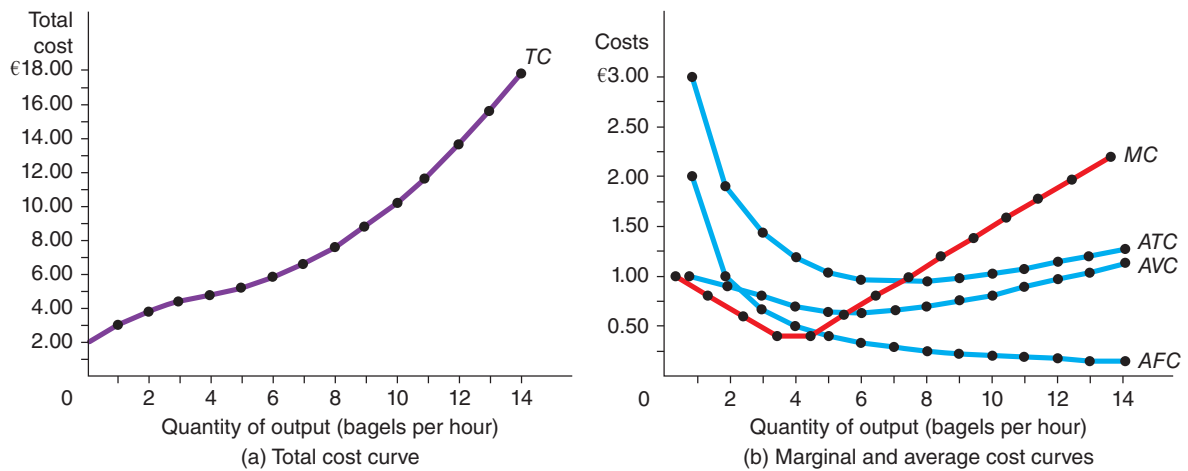
In the examples we have studied so far, the firms exhibit diminishing marginal product and, therefore, rising marginal cost at **all** levels of output. Yet actual firms are often a bit more complicated than this. In many firms, diminishing marginal product does not start to occur immediately after the first worker is hired. Depending on the production process, the second or third worker might have higher marginal product than the first because a team of workers can divide tasks and work more productively than a single worker. Such firms would first experience increasing marginal product for a while before diminishing marginal product sets in.

The table in Figure 6.4 shows the cost data for such a firm, called Berit's Bagel Bin. These data are used in the graphs. Panel (a) shows how total cost (TC) depends on the quantity produced, and panel (b) shows average total cost (ATC), average fixed cost (AFC), average variable cost (AVC) and marginal cost (MC). In the range of output from 0 to 4 bagels per hour, the firm experiences increasing marginal product, and

the marginal cost curve falls. After 5 bagels per hour, the firm starts to experience diminishing marginal product, and the marginal cost curve starts to rise. This combination of increasing then diminishing marginal product also makes the average variable cost curve U-shaped.

FIGURE 6.4**Berit's Cost Curves**

Many firms, like Berit's Bagel Bin, experience increasing marginal product before diminishing marginal product and, therefore, have cost curves shaped like those in this figure. Panel (a) shows how total cost (TC) depends on the quantity produced. Panel (b) shows how average total cost (ATC), average fixed cost (AFC), average variable cost (AVC) and marginal cost (MC) depend on the quantity produced. These curves are derived by graphing the data from the table. Notice that marginal cost and average variable cost fall for a while before starting to rise.



| Quantity of bagels (per hour) Q | Total cost (€) $TC = FC + VC$ | Fixed cost (€) FC | Variable cost (€) VC | Average fixed cost (€) $AFC = FC/Q$ | Average variable cost (€) $AVC = VC/Q$ | Average total cost (€) $ATC = TC/Q$ | Marginal cost (€) $MC = \Delta TC / \Delta Q$ |
|--------------------------------------|----------------------------------|------------------------|---------------------------|--|---|--|--|
| 0 | 2.00 | 2.00 | 0.00 | – | – | – | 1.00 |
| 1 | 3.00 | 2.00 | 1.00 | 2.00 | 1.00 | 3.00 | 0.80 |
| 2 | 3.80 | 2.00 | 1.80 | 1.00 | 0.90 | 1.90 | 0.60 |
| 3 | 4.40 | 2.00 | 2.40 | 0.67 | 0.80 | 1.47 | 0.40 |
| 4 | 4.80 | 2.00 | 2.80 | 0.50 | 0.70 | 1.20 | 0.40 |
| 5 | 5.20 | 2.00 | 3.20 | 0.40 | 0.64 | 1.04 | 0.60 |
| 6 | 5.80 | 2.00 | 3.80 | 0.33 | 0.63 | 0.96 | 0.80 |
| 7 | 6.60 | 2.00 | 4.60 | 0.29 | 0.66 | 0.95 | 1.00 |
| 8 | 7.60 | 2.00 | 5.60 | 0.25 | 0.70 | 0.95 | 1.20 |
| 9 | 8.80 | 2.00 | 6.80 | 0.22 | 0.76 | 0.98 | 1.40 |
| 10 | 10.20 | 2.00 | 8.20 | 0.20 | 0.82 | 1.02 | 1.60 |
| 11 | 11.80 | 2.00 | 9.80 | 0.18 | 0.89 | 1.07 | 1.80 |
| 12 | 13.60 | 2.00 | 11.60 | 0.17 | 0.97 | 1.14 | 2.00 |
| 13 | 15.60 | 2.00 | 13.60 | 0.15 | 1.05 | 1.20 | 2.20 |
| 14 | 17.80 | 2.00 | 15.80 | 0.14 | 1.13 | 1.27 | |

Despite these differences from our previous example, Berit's cost curves share the three properties that are most important to remember:

- Marginal cost eventually rises with the quantity of output.
- The average total cost curve is U-shaped.
- The marginal cost curve crosses the average total cost curve at the minimum of average total cost.

SELF TEST A firm's total cost function is $TC = 0.000002q^3 - 0.006q^2 + 8q$. Find the total cost if output is 1,000 units. What is the average total cost at 1,000 units? What are the average fixed costs of 1,000 units? What is the marginal cost of the 1,000th unit?

COSTS IN THE SHORT RUN AND IN THE LONG RUN

So far we have analysed costs in the short run under the assumption that some factors of production, such as the size of Paolo's factory, cannot be changed. We are now going to look at what happens when that assumption is relaxed.

The Relationship Between Short-Run and Long-Run Average Total Cost

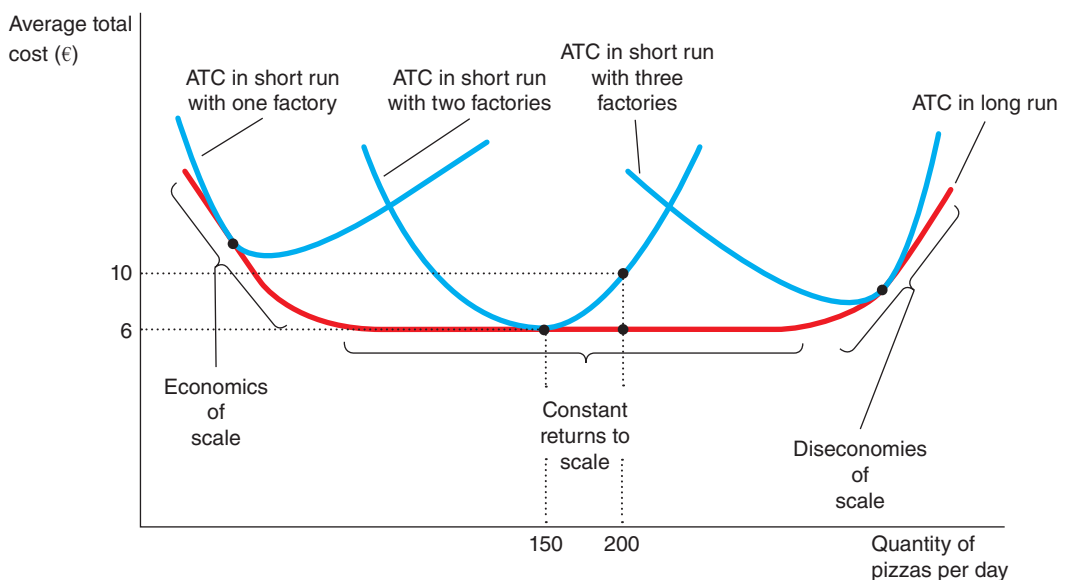
For many firms, the division of total costs between fixed and variable costs depends on the time horizon. Consider, for instance, Paolo's Pizza Factory. Over a period of a few months, Paolo cannot expand the size of his factory. The only way he can produce additional pizzas is to hire more workers at the factory he already has. The cost of Paolo's factory is, therefore, a fixed cost in the short run. By contrast, over a period of several years, Paolo can expand the size of his factory, build or buy new factories. Thus, the cost of Paolo's factories is a variable cost in the long run.

Because many decisions are fixed in the short run but variable in the long run, a firm's long-run cost curves differ from its short-run cost curves. Figure 6.5 shows an example. The figure presents three short-run average total cost curves representing the cost structures for three factories. It also presents the long-run average total cost curve. As Paolo adjusts the size of his capacity to the quantity of production, he moves along the long-run curve, and he adjusts capacity to the quantity of production.

FIGURE 6.5

Average Total Cost in the Short and Long Runs

Because fixed costs are variable in the long run, the average total cost curve in the short run differs from the average total cost curve in the long run.



This graph shows how short-run and long-run costs are related. The long-run average total cost curve is a much flatter U-shape than the short-run average total cost curve. In addition, all the short-run curves lie on or above the long-run curve. These properties arise because firms have greater flexibility in the long run. In essence, in the long run, the firm chooses which short-run curve it wants to use. But in the short run, it has to use whatever short-run curve it chose in the past.

The figure shows an example of how a change in production alters costs over different time horizons. When Paolo wants to increase production from 150 to 200 pizzas per day, he has no choice in the short run but to hire more workers with only two factories. Because of diminishing marginal product, average total cost rises from €6 to €10 per pizza. In the long run, however, Paolo can expand both his capacity and his workforce by building or acquiring a third factory, and average total cost returns to €6 per pizza.

How long does it take for a firm to get to the long run? The answer depends on the firm. It can take several years for a major manufacturing firm, such as a car company, to build a larger factory. By contrast, Paolo might be able to find new premises and expand sales in a matter of months. There is, therefore, no single answer about how long it takes a firm to adjust its production facilities.

Why is the long-run average total cost curve often U-shaped? At low levels of production, the firm benefits from increased size because it can take advantage of certain benefits such as greater specialization. Being larger it might suffer from coordination problems but these may not yet be acute. By contrast, at high levels of production, the benefits of specialization have already been realized, and coordination problems become more severe as the firm grows larger. Thus, long-run average total cost is falling at low levels of production because of increasing specialization and rising at high levels of production because of increasing coordination problems.

SELF TEST Why is the short run average cost curve u-shaped? Why is the long run average cost curve also u-shaped?

SUMMARY

To help summarize the important points we have raised so far in this chapter, Table 6.3 summarizes some of the definitions we have encountered.

TABLE 6.3 The Many Types of Cost: A Summary

| Term | Definition | Mathematical description |
|-----------------------|---|---------------------------|
| Explicit costs | Costs that require an outlay of money by the firm | – |
| Implicit costs | Costs that do not require an outlay of money by the firm | – |
| Fixed costs | Costs that do not vary with the quantity of output produced | FC |
| Variable costs | Costs that do vary with the quantity of output produced | VC |
| Total cost | The market value of all the inputs that a firm uses in production | $TC = FC + VC$ |
| Average fixed cost | Fixed costs divided by the quantity of output | $AFC = FC/Q$ |
| Average variable cost | Variable costs divided by the quantity of output | $AVC = VC/Q$ |
| Average total cost | Total cost divided by the quantity of output | $ATC = TC/Q$ |
| Marginal cost | The increase in total cost that arises from an extra unit of production | $MC = \Delta TC/\Delta Q$ |

RETURNS TO SCALE

Our analysis has shown that in the short run the firm has some options for increasing output by varying some factors of production – it can hire more workers or reorganize production processes, for example. In the long run, the firm is able to increase all factors of production and change capacity. Figure 6.5 shows three effects on production in the long run which are referred to as returns to scale.

Assume that business has been good for our pizza factory owner. Paolo has utilized his existing resources efficiently but finds that he is not able to supply the market with all that he could because of the constraints of the capacity of the factory. He knows that increasing the number of worker hours and machine hours will generate increased output up to a point, but that diminishing marginal productivity is a likelihood if factory space is held constant.

In the long run, the firm can expand all the factors of production. Our factory owner could search for a new site on which to build a new factory or could purchase an existing factory and equip it to produce pizzas. If all the factors of production are changed then the firm will be able to operate at a different scale.

Economies and Diseconomies of Scale

When a firm increases the scale of production there are three outcomes that can occur as highlighted in Figure 6.5. Assume our pizza factory owner currently employs 50 workers and 10 machines in a factory with a floor space of 1,000m² and currently produces 2,000 pizzas per day. The total cost of producing these 2,000 pizzas per day is €4,000. The average cost of each pizza is, therefore, €4,000/2,000 = €2 each.

Paolo doubles the input of all the factors of production and as a result now employs 100 workers, 20 machines and has 2,000m² of capacity. The cost of expanding the scale of production is clearly going to be higher but what happens to the average cost of production? This depends on how far total cost increases and what happens to output. If the total cost of production also doubled to €8,000 and output also doubled to 4,000, the average cost of each pizza would still be €2. The firm is said to be experiencing **constant returns to scale**. This occurs when long-run average total cost does not vary with the level of output. If, however, the total cost of production at the new scale of production increased to €6,000 and output doubled to 4,000, the average cost of each pizza would now fall to €1.50 each. In this case the firm will experience increasing returns to scale. In this situation the proportionate increase in output is greater than the proportionate increase in total cost. In such a situation the firm is said to be experiencing **economies of scale** which occur when long-run average total cost declines as output increases.

constant returns to scale the property whereby long-run average total cost stays the same as the quantity of output changes
economies of scale the property whereby long-run average total cost falls as the quantity of output increases

If the doubling of factor inputs leads to an increase in total costs (to €10,000, for example) which is greater than the increase in output (assume this is 4,000) the firm is said to experience decreasing returns to scale. In this situation the average cost per pizza would now be €2.50 each. When long-run average total cost rises as output increases, there are said to be **diseconomies of scale**.

diseconomies of scale the property whereby long-run average total cost rises as the quantity of output increases

Note that when we talk about economies of scale we are referring to **unit costs** (or average costs) – clearly if a firm increases its capacity by building a new factory and then hiring more capital and labour to work in the factory, total costs are going to rise but if the relative increase in output is greater than the relative increase in total costs as a result then the unit or average cost will fall. Note that here we are referring to the concept of scale as the proportionate increase of all factor inputs and the resultant relative increase in output. Consideration has to be taken of the price of factor inputs. In our example, the price of labour is €60 per unit and the price of capital €100 per unit. If these prices remain constant then a 50 per cent

increase in all factor inputs would increase total cost by 50 per cent. If this 50 per cent increase in inputs leads to a 75 per cent increase in output then average costs will fall.

Returns to scale can be found by using the formula:

$$\text{Returns to scale} = \frac{\% \Delta \text{ quantity of output}}{\% \Delta \text{ in the quantity of all factor inputs}}$$

Causes of Economies and Diseconomies of Scale What might cause economies or diseconomies of scale? Economies of scale often arise because higher production levels allow *specialization* among workers and increase the possibility that technology can be used, which permits each worker to become better at his or her assigned tasks. For instance, modern assembly line production for furniture may require fewer workers in relation to the technology used but still produce more furniture. If a furniture manufacturer was producing only a small quantity of output, it could not take advantage of this approach and would have higher average total cost.

In addition to reasons of more effective use of technology, firms operating at larger scale can also gain advantages by being in a position to negotiate more favourable borrowing rates or can secure lower supply costs because they are buying in larger quantities and can agree discounts from suppliers.

There are also advantages to larger firms from external sources which are to do with the growth of the industry as a whole rather than the increase in the scale of operation. Some firms are located in areas where there is a supply of skilled labour or which has a reputation which can benefit the firm. The development of local infrastructure may also give the firm advantages which all lead to lower unit costs.

CASE STUDY The Implications of Economies of Scale

Economies of scale are the advantages of large-scale production that result in lower average or unit costs. Imagine a firm which makes bricks. The current plant has a maximum capacity of 100,000 bricks per week and the total costs are €200,000 per week. The average cost for each brick, therefore, is €2. The firm sets a price of €2.20 per brick giving it a profit margin of €0.20 per brick.

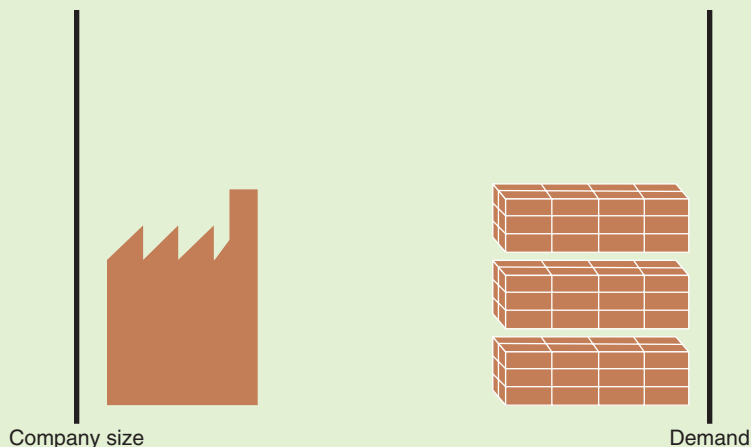
Total costs: €200,000
per week

Price per brick: €2.20

Max. capacity: 100,000
per week

Max. revenue =
€220,000 per week

Average cost per brick = €2.00



Now imagine that, in the long run, the firm expands. It doubles the size of its plant. The total costs, obviously, increase – they are now using more land and acquire more capital, as well as hiring extra labour and buying more equipment and raw materials. All of this expansion will increase the total cost. However, it is not the case that a doubling of capacity will also lead to a doubling of the cost.

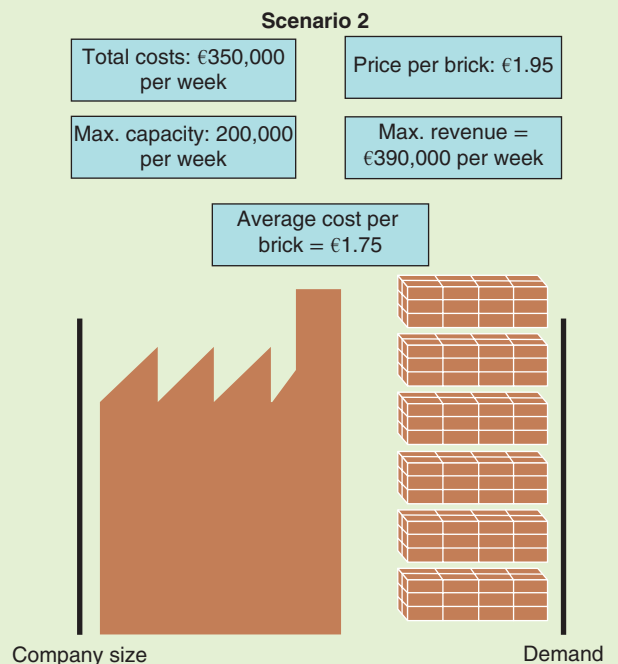
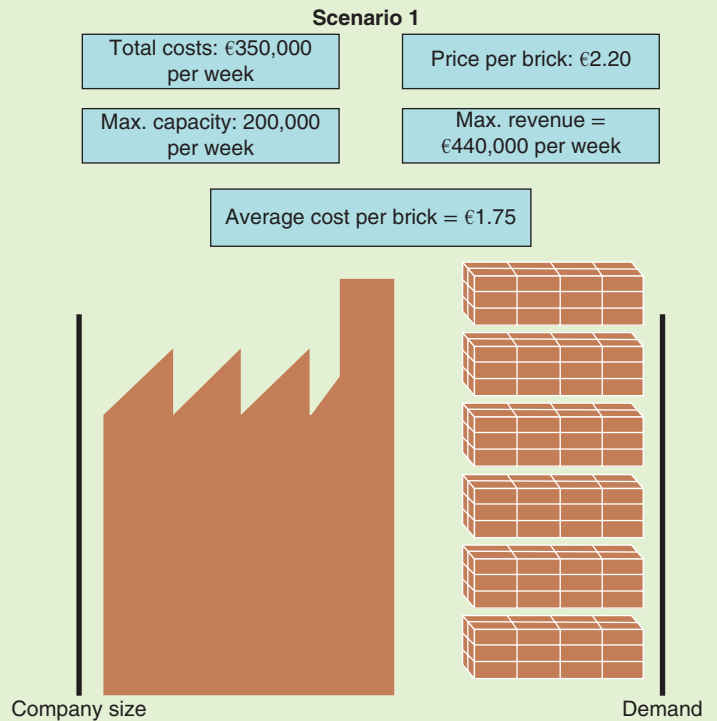
Assume TC is now €350,000 per week. The expansion of the plant means that the firm can double its output so its capacity is now 200,000 bricks per week. The percentage increase in the total costs is less than the percentage increase in output. Total costs have risen by €150,000 or 75 per cent and total output by 100 per cent, which means that the average cost per brick is now €1.75.

The firm now faces two scenarios. In Scenario 1, the firm could maintain its price at €2.20 and increase its profit margin on each brick sold from €0.20 to €0.45. Assuming it sells all the bricks it produces its revenue would increase to €440,000 per week.

In Scenario 2, the firm might choose to reduce its price to improve its competitiveness against its rivals. It could maintain its former profit margin of €0.20 and reduce the price to €1.95. In this case, if it sells all it produces its revenue would be €390,000 per week.

What the firm chooses to do would be dependent on its competitive position. If it played a dominant role in the market it might be able to increase its price and still sell all it produces. If it was in a more competitive market it might not have sold all its capacity in the first place so being able to reduce its price might mean that it can now increase sales against its rivals and increase its total revenue as a result.

Economies of scale, therefore, occur where the proportionate rise in output as a result of the expansion or growth of the firm, as defined by a rise in all the factor inputs, is greater than the proportionate rise in costs as a result of the expansion.



Diseconomies of scale can arise because of coordination and communication problems that are inherent in any large organization. The bigger the business organization, the more stretched the management team becomes, the more likely it is that communication and decision making become dysfunctional and the less effective the managers become at keeping costs down.

SELF TEST If Airbus produces 9 jets per month, its long-run total cost is €9.0 million per month. If it produces 10 jets per month, its long-run total cost is €9.5 million per month. Does Airbus exhibit economies or diseconomies of scale?

WHAT IS A COMPETITIVE MARKET?

Having looked at a firm's costs we now turn our attention to a firm's revenue. To do this we need to consider the nature of a competitive market.

The Meaning of Competition

Competition refers to a situation where there are rivals in production who allow the consumer to make a choice. A market in which there is only one supplier has no competition as consumers have no choice, but a market where there are many suppliers gives consumers the opportunity of making a choice based on quality, value for money, price and so on. We can summarize the meaning of competition in the following way:

- Where more than one firm offers the same or a similar product there is competition.
- Competition can also manifest itself where substitutes exist: for example, gas and electricity are separate markets but there is the opportunity for consumers to substitute gas cookers for electric ones and so some element of competition exists.
- The closer the degree of substitutability the greater will be the competition that exists.
- Firms may influence the level of competition through the way they build relationships with consumers, encourage purchasing habits, provide levels of customer service and after sales service, and so on.

At the extreme end of the competitive scale is a *perfectly competitive market*. A competitive market has a number of characteristics:

- There are many buyers and many sellers in the market.
- The goods offered by the various sellers are largely the same (if identical the goods are described as being 'homogenous').
- Firms have to accept the price determined by the market as a whole – no individual firm can influence supply and thus price. Firms are referred to as being 'price takers'. Each seller can sell all he wants at the going price, he has little reason to charge less, and if he charges more buyers will go elsewhere.
- There are no restrictions (called barriers to entry) on firms entering or exiting a market. Any firm is free to set up and conduct business and equally there are no reasons why they cannot just close down and leave the industry.
- There is a high degree of information available to buyers and sellers in the market.

There are few examples of markets where all these assumptions hold all the time. However, this does not mean that the model of perfectly competitive markets is not of value. It provides a benchmark to analyse how firms would behave **if** these assumptions held. We can then look at how firms' behaviour would differ if some of these assumptions were relaxed.

An example of a market which is close to being a perfect market is the market for milk. No single buyer of milk can influence the price of milk because each buyer purchases a small amount relative to the size of the market. Similarly, each seller of milk has limited control over the price because many other sellers are offering milk that is essentially identical (although even in this market firms try and make their product appear different by offering skimmed milk, semi-skimmed milk, full cream milk, flavoured milk and so on).

Anyone can decide to start a dairy farm, and if any existing dairy farmer decides to leave the dairy business then they can. Information about the milk industry is widespread and so both consumers and producers are able to make informed decisions.

The Revenue of a Competitive Firm

A firm in a competitive market, like most other firms in the economy, tries to maximize profit, which equals total revenue minus total cost. To see how it does this, we first consider the revenue of a competitive firm. To keep matters concrete, let's consider a specific firm: the Grundy Family Dairy Farm.

The Grundy Farm produces a quantity of milk Q and sells each unit at the market price P . The farm's total revenue is $P \times Q$. For example, if a litre of milk sells for €6 and the farm sells 1,000 litres, its total revenue is €6,000.

Because the Grundy Farm is small compared with the world market for milk, it takes the price as given by market conditions. This means, in particular, that the price of milk does not depend on the quantity of output that the Grundy Farm produces and sells. If the Grundy's double the amount of milk they produce, the price of milk remains the same and their total revenue doubles. As a result, total revenue is proportional to the amount of output.

Table 6.4 shows the revenue for the Grundy Family Dairy Farm. The first two columns show the amount of output the farm produces and the price at which it sells its output. The third column is the farm's total revenue. The table assumes that the price of milk is €6 a litre, so total revenue is simply €6 times the number of litres. (The price of milk per litre is much less than this across Europe but we have made this assumption to keep the numbers simple and to illustrate the principle).

TABLE 6.4 Total, Average and Marginal Revenue for a Competitive Firm

| Quantity (Q) | Price (€) (P) | Total revenue (€) ($TR = P \times Q$) | Average revenue (€) ($AR = TR/Q$) | Marginal revenue (€) ($MR = \Delta TR/\Delta Q$) |
|---------------------|----------------------|---|---|--|
| 1 litre | 6 | 6 | 6 | |
| 2 | 6 | 12 | 6 | 6 |
| 3 | 6 | 18 | 6 | 6 |
| 4 | 6 | 24 | 6 | 6 |
| 5 | 6 | 30 | 6 | 6 |
| 6 | 6 | 36 | 6 | 6 |
| 7 | 6 | 42 | 6 | 6 |
| 8 | 6 | 48 | 6 | 6 |

As we did when looking at a firm's costs, consider these two questions:

- How much revenue does the farm receive for the typical litre of milk?
- How much additional revenue does the farm receive if it increases production of milk by 1 litre?

The last two columns in Table 6.4 answer these questions.

The fourth column in the table shows **average revenue**, which is total revenue (from the third column) divided by the amount of output (from the first column).

$$AR = \frac{TR}{Q}$$

Average revenue tells us how much revenue a firm receives for the typical unit sold. In Table 6.4, you can see that average revenue equals €6, the price of a litre of milk. This illustrates a general lesson that applies not only to competitive firms but to other firms as well. Total revenue is the price times the quantity ($P \times Q$), and average revenue is total revenue ($P \times Q$) divided by the quantity (Q). Therefore, *for all firms, average revenue equals the price of the good.*

average revenue total revenue divided by the quantity sold

The fifth column shows **marginal revenue**, which is the change in total revenue from the sale of each additional unit of output.

$$MR = \frac{\Delta TR}{\Delta Q}$$

Using calculus:

$$\frac{dTR}{dQ}$$

marginal revenue the change in total revenue from an additional unit sold

The sale of one more litre of milk adds €6 to total revenue, therefore the marginal revenue is also €6. In Table 6.1, marginal revenue equals €6, the price of a litre of milk. This result illustrates a lesson that applies only to competitive firms. Total revenue is $P \times Q$, and P is fixed for a competitive firm. Therefore, when Q rises by 1 unit, total revenue rises by P euros. For competitive firms, marginal revenue equals the price of the good.

SELF TEST When a competitive firm doubles the amount it sells, what happens to the price of its output and its total revenue?

Total Revenue, Total Cost and Profit

It is conceivable that Paolo or a dairy farmer started in business because of a desire to provide the world with pizza or milk or, perhaps, out of love for the pizza business or farming. More likely, they started in business to make money. Economists often use the assumption that the goal of a firm is to maximize profit. Whilst the extent to which this assumption holds in the real world has been questioned, it is a useful starting point for our analysis.

What is a firm's profit? **Profit** is the difference between total revenue and total cost.

$$\text{Profit} = \text{Total revenue} - \text{Total cost}$$

We can express this in the formula:

$$\pi = TR - TC$$

where π represents profit.

We will assume that a firm's objective is to make profit as large as possible – in other words, they want to maximize profit.

Economic Profit versus Accounting Profit

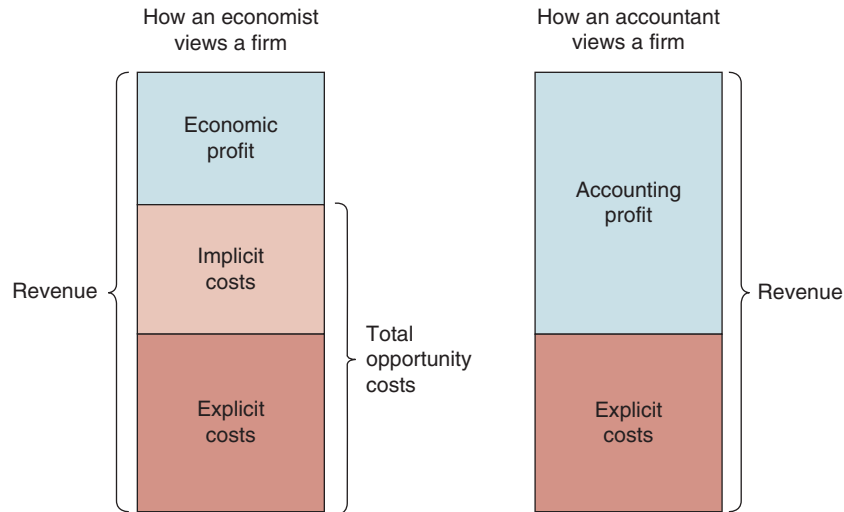
Economists and accountants measure costs differently, they also measure profit differently. An economist measures a firm's **economic profit** as the firm's total revenue minus all the opportunity costs (explicit and implicit) of producing the goods and services sold. An accountant measures the firm's **accounting profit** as the firm's total revenue minus only the firm's explicit costs.

economic profit total revenue minus total cost, including both explicit and implicit costs
accounting profit total revenue minus total explicit cost

Figure 6.6 summarizes this difference. Notice that because the accountant ignores the implicit costs, accounting profit is usually larger than economic profit. For a business to be profitable from an economist's standpoint, total revenue must cover all the opportunity costs, both explicit and implicit.

FIGURE 6.6**Economists versus Accountants**

Economists include all opportunity costs when analysing a firm, whereas accountants measure only explicit costs. Therefore, economic profit is smaller than accounting profit.



PROFIT MAXIMIZATION AND THE COMPETITIVE FIRM'S SUPPLY CURVE

We are making an assumption that one goal of a competitive firm is to maximize profit. We are now ready to examine how the firm maximizes profit and how that decision leads to its supply curve.

A Simple Example of Profit Maximization

Let's begin our analysis of the firm's supply decision with the example in Table 6.5 using the Grundy Family Dairy Farm. In the first column of the table is the number of litres of milk the farm produces. The second column shows the farm's total revenue, which is €6 times the number of litres. The third column shows the farm's total cost. Total cost includes fixed costs, which are €3 in this example, and variable costs, which depend on the quantity produced.

The fourth column shows the farm's profit, which is computed by subtracting total cost from total revenue. If the farm produces nothing, it has a loss of €3. If it produces 1 litre, it has a profit of €1. If it produces 2 litres, it has a profit of €4 and so on. To maximize profit, the Grundy Farm chooses the quantity that makes profit as large as possible. In this example, profit is maximized when the farm produces 4 or 5 litres of milk, when the profit is €7.

There is another way to look at the Grundy Farm's decision: the Grundy's can find the profit-maximizing quantity by comparing the marginal revenue and marginal cost from each unit produced. The fifth and sixth columns in Table 6.5 compute marginal revenue and marginal cost from the changes in total revenue and total cost, and the last column shows the change in profit for each additional litre produced. The first litre of milk the farm produces has a marginal revenue of €6 and a marginal cost of €2; hence, producing that litre adds €4 to profit (from –€3 to €1). The second litre produced has a marginal revenue of €6 and a marginal cost of €3, so that additional litre adds €3 to profit (from €1 to €4). As long as marginal revenue exceeds marginal cost, increasing the quantity produced **adds** to profit. Since additional production of milk adds to profit

TABLE 6.5 Profit Maximization: A Numerical Example

| Quantity (Q) | Total revenue (€) (TR) | Total cost (€) (TC) | Profit (€) ($TR - TC$) | Marginal revenue (€) ($MR = \Delta TR / \Delta Q$) | Marginal cost (€) ($MC = \Delta TC / \Delta Q$) | Change in profit (€) ($MR - MC$) |
|---------------------|----------------------------------|----------------------------|-----------------------------|--|---|--|
| 0 litres | 0 | 3 | -3 | | | |
| 1 | 6 | 5 | 1 | 6 | 2 | 4 |
| 2 | 12 | 8 | 4 | 6 | 3 | 3 |
| 3 | 18 | 12 | 6 | 6 | 4 | 2 |
| 4 | 24 | 17 | 7 | 6 | 5 | 1 |
| 5 | 30 | 23 | 7 | 6 | 6 | 0 |
| 6 | 36 | 30 | 6 | 6 | 7 | -1 |
| 7 | 42 | 38 | 4 | 6 | 8 | -2 |
| 8 | 48 | 47 | 1 | 6 | 9 | -3 |

it is worth the Grundy Farm producing this extra milk. Once the Grundy Farm has reached 5 litres of milk, however, the situation is very different. The sixth litre would have marginal revenue of €6 and marginal cost of €7, so producing it would reduce profit by €1 (from €7 to €6). It does not make sense for the Grundy Farm to produce the sixth litre and so, as a result, there is no incentive for the Grundy's to produce beyond 5 litres.

This is another example of thinking at the margin. If marginal revenue is greater than marginal cost – as it is at 1, 2 or 3 litres – it is worth the Grundy's increasing the production of milk; even making a decision to increase production from the third to the fourth litre is worth doing since the MC and MR are the same at this point and as a result the profit level is neither increased nor decreased. If marginal revenue is less than marginal cost, as it is at 5, 6 or 7 litres, the Grundy's should decrease production. If the Grundy's think at the margin and make incremental adjustments to the level of production, they are naturally led to produce the profit-maximizing quantity. The profit maximizing output occurs, therefore, at the output where $MR = MC$.

Thinking at the margin means that farmers such as the Grundy's will consider the marginal cost and marginal revenue of an additional cow when considering profit.



Normal and Abnormal Profit

In the analysis that follows we make an important assumption related to our earlier discussion of the meaning of economic profit. We know that profit equals total revenue minus total cost, and that total cost includes the opportunity cost of the time and money that the firm owners devote to the business. A firm's revenue must compensate the owners for the time and money that they expend to keep their business going, which is sometimes referred to as normal profit or zero-profit equilibrium.

Consider an example. Suppose that a farmer had to invest €1 million to open his farm, which otherwise he could have deposited in a bank to earn €50,000 a year in interest. In addition, he had to give up another job that would have paid him €30,000 a year. Then the farmer's opportunity cost of farming includes both the interest he could have earned and the forgone wages – a total of €80,000. This sum must be calculated as part of the farmer's total costs. This is referred to as **normal profit** – the minimum amount required to keep factor inputs in their current use. Even if his profit is driven to zero, his revenue from farming compensates him for these opportunity costs.

normal profit the minimum amount required to keep factors of production in their current use

Because of the different way accountants and economists measure profit, at the zero-profit equilibrium, economic profit is zero, but accounting profit is positive. Our farmer's accountant, for instance, would conclude that the farmer earned an accounting profit of €80,000, which is enough to keep the farmer in business. In the short run, as we shall see, profit can be above zero or normal profit which is referred to as **abnormal profit**.

abnormal profit the profit over and above normal profit

If firms are making abnormal profit then there is an incentive for other firms to enter the market to take advantage of the profits that exist and this creates a dynamic which moves the market to equilibrium. We will look at these dynamic next.

The Marginal Cost Curve and the Firm's Supply Decision

To extend this analysis of profit maximization, consider the cost curves in Figure 6.7. These cost curves have the three features discussed earlier in this chapter:

- the marginal cost curve (MC) is upwards sloping
- the average total cost curve (ATC) is U-shaped and
- the marginal cost curve crosses the average total cost curve at the minimum of average total cost.

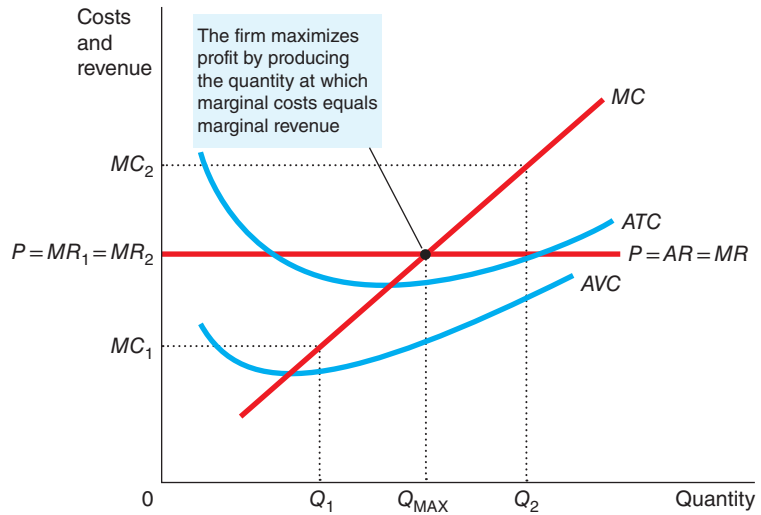
The figure also shows a horizontal line at the market price (P). The price line is horizontal because the firm is a price taker: the price of the firm's output is the same, regardless of the quantity that the firm decides to produce. Keep in mind that, for a competitive firm, the firm's price equals both its average revenue (AR) and its marginal revenue (MR).

We can use Figure 6.7 to find the quantity of output that maximizes profit. Imagine that the firm is producing at Q_1 . At this level of output, marginal revenue is greater than marginal cost. That is, if the firm raised its level of production and sales by 1 unit, the additional revenue (MR_1) would exceed the additional costs (MC_1). Profit, which equals **total** revenue minus **total** cost, would increase. Hence, if marginal revenue is greater than marginal cost, as it is at Q_1 , it is worth the firm producing this output because it can increase profit.

A similar argument applies when output is at Q_2 . In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved (MC_2) would exceed the revenue lost (MR_2). Therefore, if marginal revenue is less than marginal cost, as it is at Q_2 , it is worth the firm cutting back production because the firm can increase profit.

FIGURE 6.7**Profit Maximization for a Competitive Firm**

This figure shows the marginal cost curve (MC), the average total cost curve (ATC) and the average variable cost curve (AVC). It also shows the market price (P), which equals marginal revenue (MR) and average revenue (AR). At the quantity Q_1 , marginal revenue MR_1 exceeds marginal cost MC_1 , so raising production increases profit. At the quantity Q_2 , marginal cost MC_2 is above marginal revenue MR_2 , so reducing production increases profit. The profit maximizing quantity Q_{MAX} is found where the horizontal price line intersects the marginal cost curve.



Where do these marginal adjustments to level of production end? Regardless of whether the firm begins with production at a low level (such as Q_1) or at a high level (such as Q_2), there is an incentive for the firm to adjust its output decisions until the quantity produced reaches Q_{MAX} . At any other output level, there is an incentive for the firm to either increase or cut back output and increase profit. This analysis shows a general rule for profit maximization: at the profit-maximizing level of output, marginal revenue and marginal cost are exactly equal.

We can now see how the supply curve is derived. Because a competitive firm is a price taker, its marginal revenue equals the market price. For any given price, the competitive firm's profit-maximizing quantity of output is found by looking at the intersection of the price with the marginal cost curve. In Figure 6.7, that quantity of output is Q_{MAX} .

FIGURE 6.8**Marginal Cost as the Competitive Firm's Supply Curve**

An increase in the price from P_1 to P_2 leads to an increase in the firm's profit-maximizing quantity from Q_1 to Q_2 . Because the marginal cost curve shows the quantity supplied by the firm at any given price, it is the firm's supply curve.

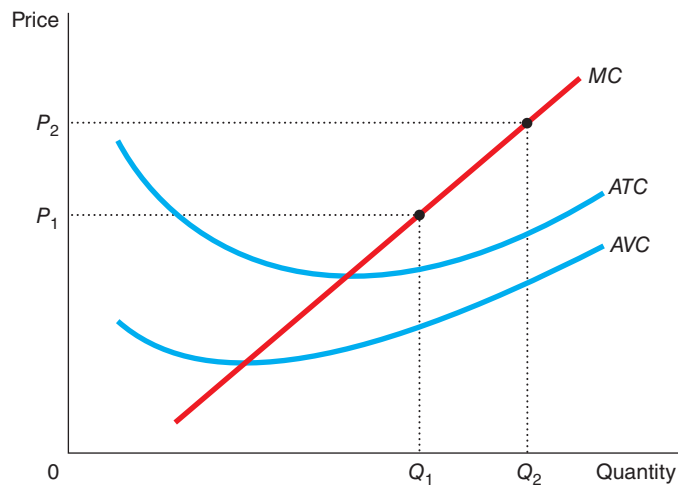


Figure 6.8 shows how a competitive firm responds to an increase in the price which may have been caused by a change in global market conditions. Remember that competitive firms are price takers and have to accept the market price for their product. Prices of commodities such as grain, metals, sugar, cotton, coffee, pork bellies, oil and so on are set by organized international markets and so the individual

firm has no power to influence price. When the price is P_1 , the firm produces quantity Q_1 , the quantity that equates marginal cost to the price. Assume that an outbreak of bovine spongiform encephalopathy (BSE) results in the need to slaughter a large proportion of dairy cattle and as a result there is a shortage of milk on the market. When the price rises to P_2 , the firm finds that marginal revenue is now higher than marginal cost at the previous level of output, so existing farmers look to increase production. The new profit-maximizing quantity is Q_2 , at which marginal cost equals the new higher price. In essence, because the firm's marginal cost curve determines the quantity of the good the firm is willing to supply at any price, it is the competitive firm's supply curve.

The Firm's Short-Run Decision to Shut Down

So far we have been analysing the question of how much a competitive firm will produce. In some circumstances, however, the firm will decide to shut down and not produce anything at all.

Here we should distinguish between a temporary shutdown of a firm and the permanent exit of a firm from the market. A *shutdown* refers to a short-run decision not to produce anything during a specific period of time because of current market conditions. *Exit* refers to a long-run decision to leave the market. The short-run and long-run decisions differ because most firms cannot avoid their fixed costs in the short run but can do so in the long run. That is, a firm that shuts down temporarily still has to pay its fixed costs, whereas a firm that exits the market saves both its fixed and its variable costs.

For example, consider the production decision that a farmer faces. The cost of a milking parlour is one of the farmer's fixed costs. If the farmer decides not to milk his herd for some period of time, the parlour still represents a cost he cannot recover. When making the short-run decision whether to shut down for this period, the fixed cost of the parlour is said to be a *sunk cost*. By contrast, if the farmer decides to leave dairy farming altogether, he can sell the parlour along with the rest of the farm. When making the long-run decision whether to exit the market, the cost of the parlour is not sunk. (We return to the issue of sunk costs shortly.)

Now let's consider what determines a firm's shutdown decision. If the firm shuts down, it loses all revenue from the sale of its product. At the same time, it saves the variable costs of making its product (but must still pay the fixed costs). Thus, the firm shuts down if the revenue that it would get from producing is less than its variable costs of production; it is simply not worth producing a product which costs more to produce than the revenue generated by its sale. Doing so would reduce profit or make any existing losses even greater.

A little bit of mathematics can make this shutdown criterion more useful. If TR stands for total revenue and VC stands for variable costs, then the firm's decision can be written as:

$$\text{Shut down if } TR < VC$$

The firm shuts down if total revenue is less than variable cost. By dividing both sides of this inequality by the quantity Q , we can write it as:

$$\text{Shut down if } TR/Q < VC/Q$$

Notice that this can be further simplified. TR/Q is total revenue divided by quantity, which is average revenue. As we discussed previously, average revenue for any firm is simply the good's price P . Similarly, VC/Q is average variable cost AVC . Therefore, the firm's shutdown criterion is:

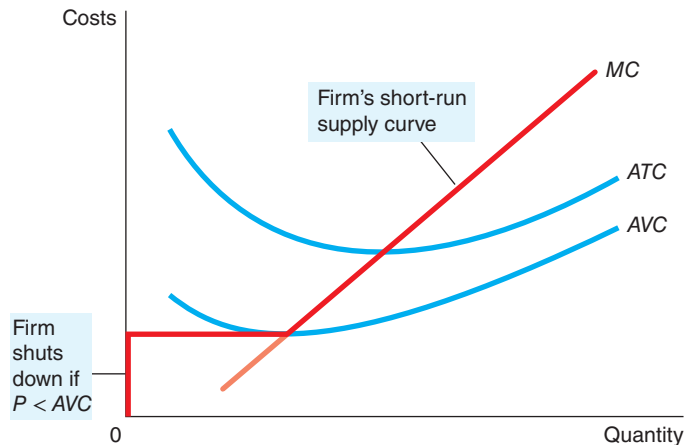
$$\text{Shut down if } P < AVC$$

That is, a firm chooses to shut down if the price of the good is less than the average variable cost of production. This criterion is intuitive: when choosing to produce, the firm compares the price it receives for the typical unit to the average variable cost that it must incur to produce the typical unit. If the price doesn't cover the average variable cost, the firm is better off stopping production altogether. The firm can re-open in the future if conditions change so that price exceeds average variable cost.

We now have a full description of a competitive firm's profit-maximizing strategy. If the firm produces anything, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average variable cost at that quantity, the firm is better off shutting down and not producing anything. These results are illustrated in Figure 6.9. The competitive firm's short-run supply curve is the portion of its marginal cost curve that lies above average variable cost.

FIGURE 6.9**The Competitive Firm's Short-Run Supply Curve**

In the short run, the competitive firm's supply curve is its marginal cost curve (MC) above average variable cost (AVC). If the price falls below average variable cost, the firm is better off shutting down.



Spilt Milk and other Sunk Costs

Sometime in your life you have probably been told, 'Don't cry over spilt milk', or 'Let bygones be bygones.' These adages hold a deep truth about rational decision making. Economists say that a cost is a **sunk cost** when it has already been committed and cannot be recovered. In a sense, a sunk cost is the opposite of an opportunity cost: an opportunity cost is what you have to give up if you choose to do one thing instead of another, whereas a sunk cost cannot be avoided, regardless of the choices you make. Because nothing can be done about sunk costs, you can ignore them when making decisions about various aspects of life, including business strategy.

sunk cost a cost that has already been committed and cannot be recovered

Our analysis of the firm's shutdown decision is one example of the importance of recognizing sunk costs. We assume that the firm cannot recover its fixed costs by temporarily stopping production. As a result, the firm's fixed costs are sunk in the short run, and the firm can safely ignore these costs when deciding how much to produce. The firm's short-run supply curve is the part of the marginal cost curve that lies above average variable cost, and the size of the fixed cost does not matter for this supply decision.

The Firm's Long-Run Decision to Exit or Enter a Market

The firm's long-run decision to exit the market is similar to its shutdown decision. If the firm exits, it again will lose all revenue from the sale of its product, but now it saves on both fixed and variable costs of production. Thus, the firm exits the market if the revenue it would get from producing is less than its total costs.

We can again make this criterion more useful by writing it mathematically. If TR stands for total revenue and TC stands for total cost, then the firm's criterion can be written as:

$$\text{Exit if } TR < TC$$

The firm exits if total revenue is less than total cost. By dividing both sides of this inequality by quantity Q , we can write it as:

$$\text{Exit if } TR/Q < TC/Q$$

We can simplify this further by noting that TR/Q is average revenue, which equals the price P , and that TC/Q is average total cost ATC . Therefore, the firm's exit criterion is:

$$\text{Exit if } P < ATC$$

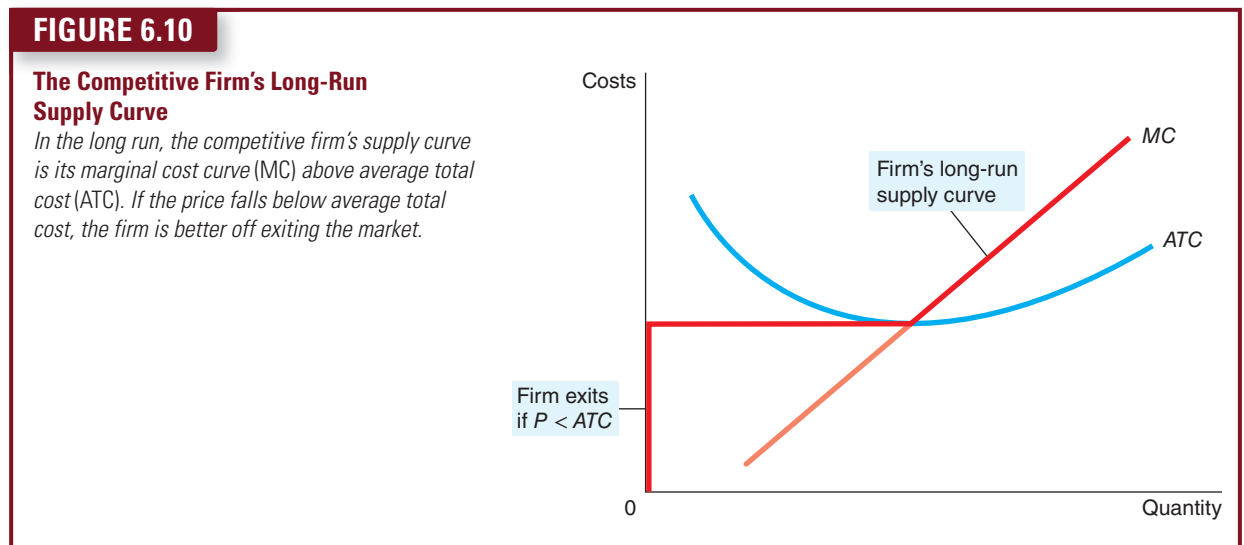
That is, a firm chooses to exit if the price of the good is less than the average total cost of production.

A parallel analysis applies to an entrepreneur who is considering starting a firm. The firm will enter the market if such an action would be profitable, which occurs if the price of the good exceeds the average total cost of production. The entry criterion is:

$$\text{Enter if } P > ATC$$

The criterion for entry is exactly the opposite of the criterion for exit.

We can now describe a competitive firm's long-run profit-maximizing strategy. If the firm is in the market, it produces the quantity at which marginal cost equals the price of the good. Yet if the price is less than average total cost at that quantity, the firm chooses to exit (or not enter) the market. These results are illustrated in Figure 6.10. The competitive firm's long-run supply curve is the portion of its marginal cost curve that lies above average total cost.



Measuring Profit in Our Graph for the Competitive Firm

As we analyse exit and entry, it is useful to be able to analyse the firm's profit in more detail. Recall that profit equals total revenue (TR) minus total cost (TC):

$$\text{Profit} = TR - TC$$

We can rewrite this definition by multiplying and dividing the right-hand side by Q :

$$\text{Profit} = (TR/Q - TC/Q) \times Q$$

But note that TR/Q is average revenue, which is the price P , and TC/Q is average total cost ATC . Therefore:

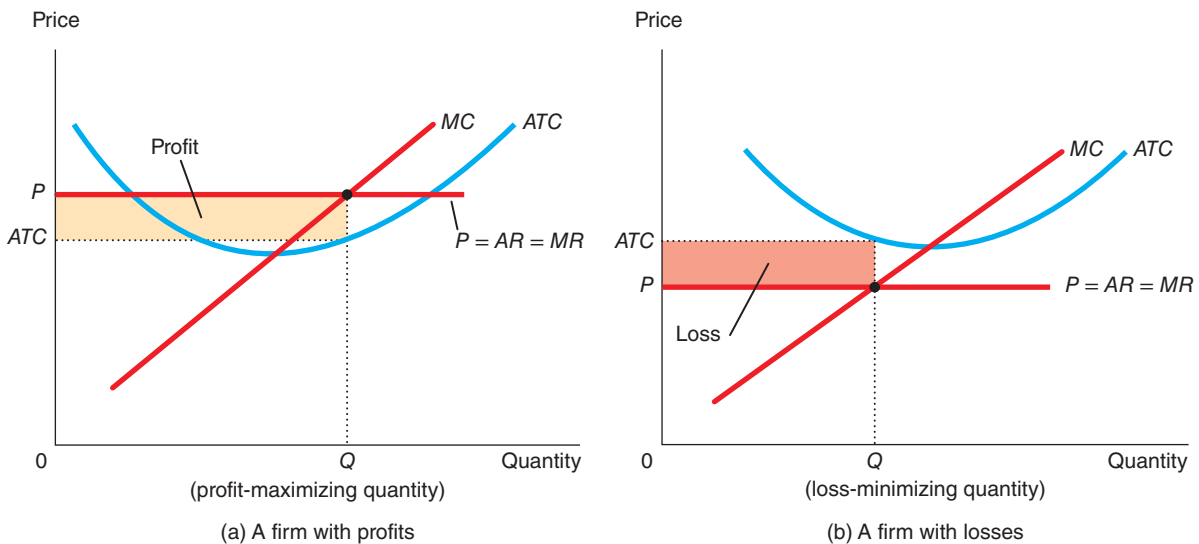
$$\text{Profit} = (P - ATC) \times Q$$

This way of expressing the firm's profit allows us to measure profit in our graphs.

Panel (a) of Figure 6.11 shows a firm earning positive profit. As we have already discussed, the firm maximizes profit by producing the quantity at which price equals marginal cost. Now look at the shaded rectangle. The height of the rectangle is $P - ATC$, the difference between price and average total cost. The width of the rectangle is Q , the quantity produced. Therefore, the area of the rectangle is $(P - ATC) \times Q$, which is the firm's profit.

FIGURE 6.11**Profit as the Area Between Price and Average Total Cost**

The area of the shaded box between price and average total cost represents the firm's profit. The height of this box is price minus average total cost ($P - ATC$), and the width of the box is the quantity of output (Q). In panel (a), price is above average total cost, so the firm has positive profit. In panel (b), price is less than average total cost, so the firm has losses.



Similarly, panel (b) of this figure shows a firm with losses (negative profit). In this case, maximizing profit means minimizing losses, a task accomplished once again by producing the quantity at which price equals marginal cost. Now consider the shaded rectangle. The height of the rectangle is $ATC - P$, and the width is Q . The area is $(ATC - P) \times Q$, which is the firm's loss. Because a firm in this situation is not making enough revenue to cover its average total cost, the firm would choose to exit the market.

SELF TEST How does the price faced by a profit-maximizing competitive firm compare to its marginal cost? Explain.

- When does a profit-maximizing competitive firm decide to shut down? When does a profit-maximizing competitive firm decide to exit a market?

THE SUPPLY CURVE IN A COMPETITIVE MARKET

Now that we have examined the supply decision of a single firm, we can discuss the supply curve for a market. There are two cases to consider. First, we examine a market with a fixed number of firms. Second, we examine a market in which the number of firms can change as old firms exit the market and new firms enter. Both cases are important, for each applies over a specific time horizon. Over short periods of time it is often difficult for firms to enter and exit, so the assumption of a fixed number of firms is appropriate. But over long periods of time, the number of firms can adjust to changing market conditions.

The Short Run: Market Supply with a Fixed Number of Firms

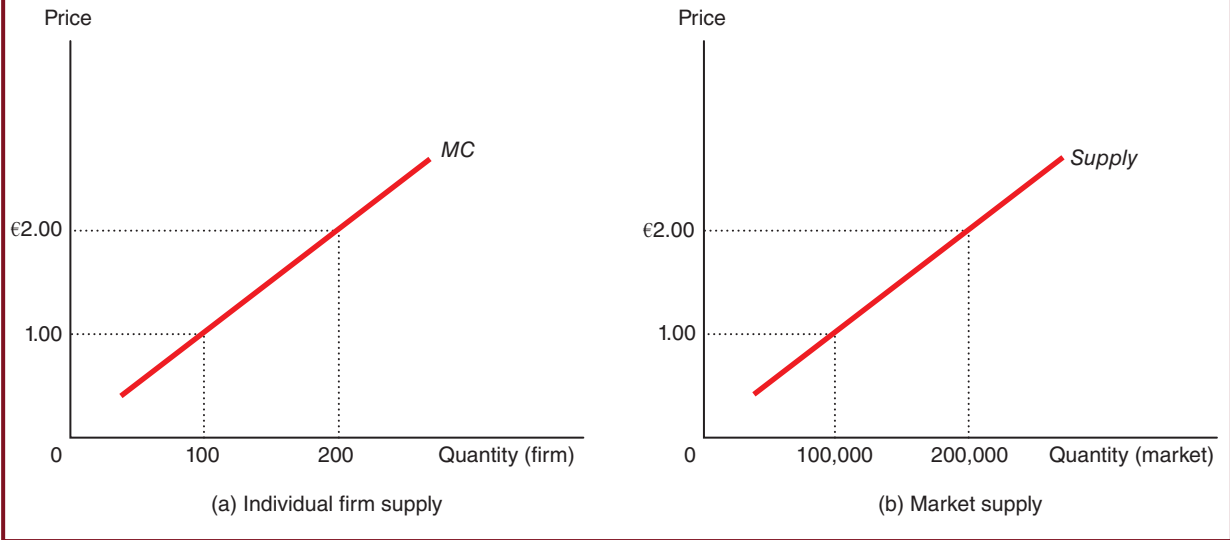
Consider first a market with 1,000 identical firms. For any given price, each firm supplies a quantity of output so that its marginal cost equals the price, as shown in panel (a) of Figure 6.12. That is, as long as price is above average variable cost, each firm's marginal cost curve is its supply curve. The quantity of output supplied to the market equals the sum of the quantities supplied by each of the 1,000 individual firms.

Thus, to derive the market supply curve, we add the quantity supplied by each firm in the market. As panel (b) of Figure 6.12 shows, because the firms are identical, the quantity supplied to the market is 1,000 times the quantity supplied by each firm.

FIGURE 6.12

Market Supply With a Fixed Number of Firms

When the number of firms in the market is fixed, the market supply curve, shown in panel (b), reflects the individual firms' marginal cost curves, shown in panel (a). Here, in a market of 1,000 firms, the quantity of output supplied to the market is 1,000 times the quantity supplied by each firm.



The Long Run: Market Supply with Entry and Exit

Now consider what happens if firms are able to enter or exit the market. Let's suppose that everyone has access to the same technology for producing the good and access to the same markets to buy the inputs into production. Therefore, all firms and all potential firms have the same cost curves.

Decisions about entry and exit in a market of this type depend on the incentives facing the owners of existing firms and the entrepreneurs who could start new firms. If firms already in the market are profitable, then new firms will have an incentive to enter the market. This entry will expand the number of firms, increase the quantity of the good supplied, and drive down prices and profits. Conversely, if firms in the market are making losses, then some existing firms will exit the market. Their exit will reduce the number of firms, decrease the quantity of the good supplied, and drive up prices and profits. At the end of this process of entry and exit, firms that remain in the market must be making zero economic profit. Recall that we can write a firm's profits as:

$$\text{Profit} = (P - ATC) \times Q$$

This equation shows that an operating firm has zero profit if, and only if, the price of the good equals the average total cost of producing that good. If price is above average total cost, profit is positive, which encourages new firms to enter. If price is less than average total cost, profit is negative, which encourages some firms to exit. The process of entry and exit ends only when price and average total cost are driven to equality.

This analysis has a surprising implication. We noted earlier in the chapter that competitive firms produce so that price equals marginal cost. We just noted that free entry and exit forces price to equal average total cost. But if price is to equal both marginal cost and average total cost, these two measures of cost must equal each other. Marginal cost and average total cost are equal, however, only when the firm is operating

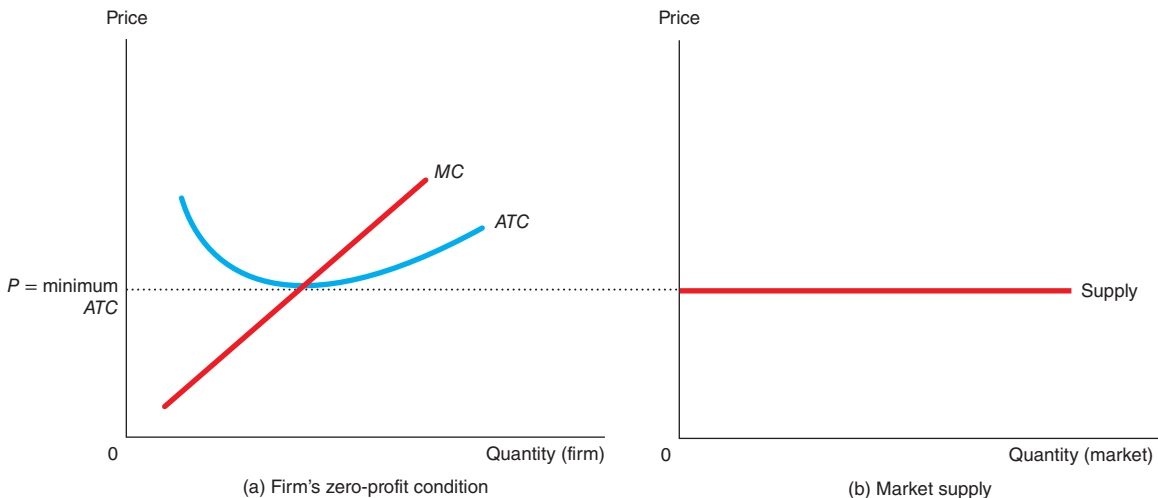
at the minimum of average total cost. Recall from earlier in this chapter that the level of production with lowest average total cost is called the firm's efficient scale. Therefore, the long-run equilibrium of a competitive market with free entry and exit must have firms operating at their efficient scale.

Panel (a) of Figure 6.13 shows a firm in such a long-run equilibrium. In this figure, price P equals marginal cost MC , so the firm is profit-maximizing. Price also equals average total cost ATC , so profits are zero. New firms have no incentive to enter the market, and existing firms have no incentive to leave the market.

From this analysis of firm behaviour, we can determine the long-run supply curve for the market. In a market with free entry and exit, there is only one price consistent with zero profit – the minimum of average total cost. As a result, the long-run market supply curve must be horizontal at this price, as in panel (b) of Figure 6.13. Any price above this level would generate profit, leading to entry and an increase in the total quantity supplied. Any price below this level would generate losses, leading to exit and a decrease in the total quantity supplied. Eventually, the number of firms in the market adjusts so that price equals the minimum of average total cost, and there are enough firms to satisfy all the demand at this price.

FIGURE 6.13**Market Supply With Entry and Exit**

Firms will enter or exit the market until profit is driven to zero. Thus, in the long run, price equals the minimum of average total cost, as shown in panel (a). The number of firms adjusts to ensure that all demand is satisfied at this price. The long-run market supply curve is horizontal at this price, as shown in panel (b).

**A Shift in Demand in the Short Run and Long Run**

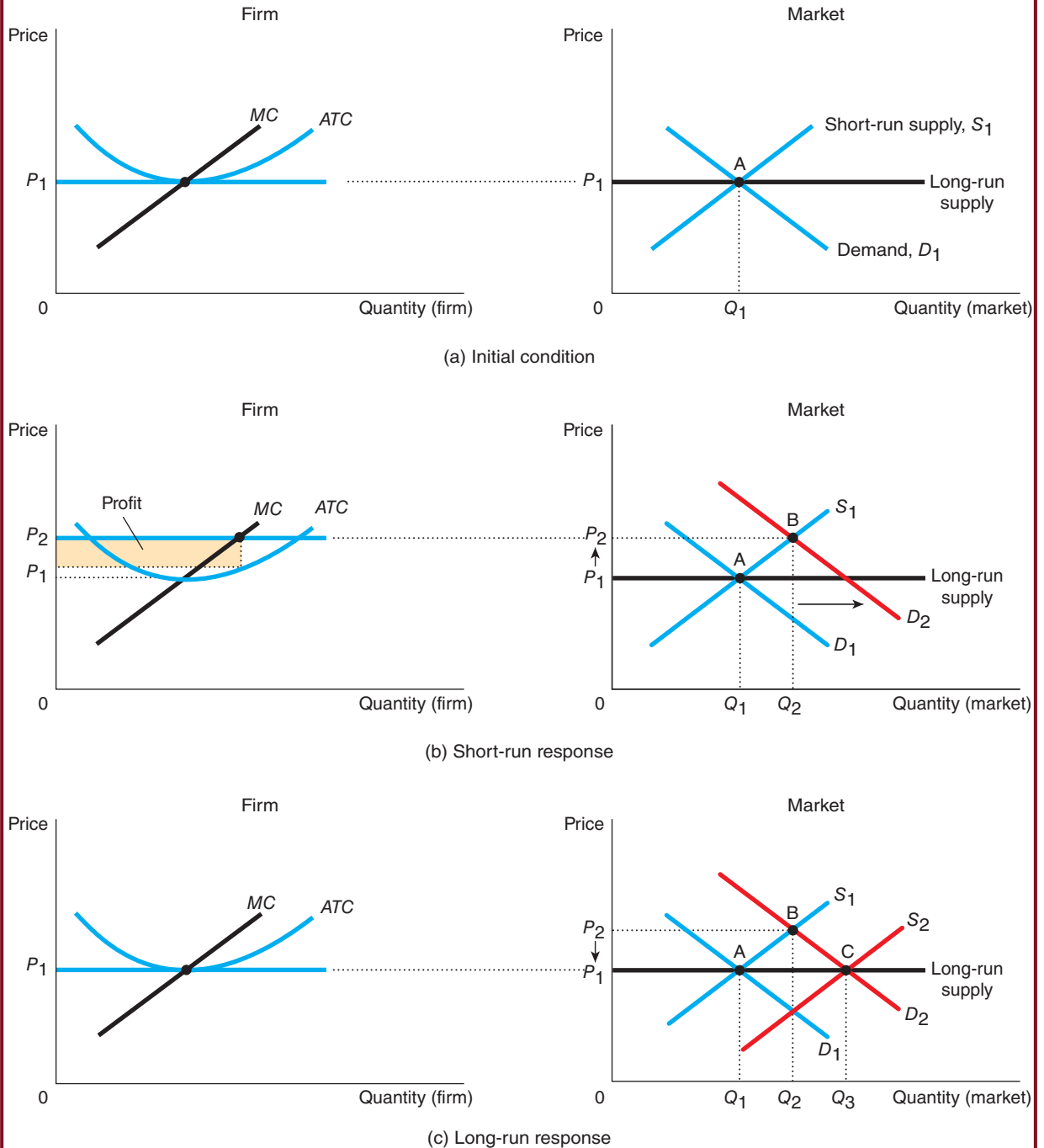
Because firms can enter and exit a market in the long run but not in the short run, the response of a market to a change in demand depends on the time horizon. To see this, let's trace the effects of a shift in demand. This analysis will show how a market responds over time, and it will show how entry and exit drive a market to its long-run equilibrium.

Suppose the market for milk begins in long-run equilibrium. Firms are earning zero profit, so price equals the minimum of average total cost. Panel (a) of Figure 6.14 shows the situation. The long-run equilibrium is point A, the quantity sold in the market is Q_1 and the price is P_1 .

Now suppose scientists discover that milk has miraculous health benefits. As a result, the demand curve for milk shifts outward from D_1 to D_2 , as in panel (b). The short-run equilibrium moves from point A to point B; as a result, the quantity rises from Q_1 to Q_2 and the price rises from P_1 to P_2 . All of the existing firms respond to the higher price by raising the amount produced. Because each firm's supply curve reflects its marginal cost curve, how much they each increase production is determined by the marginal cost curve. In the new short-run equilibrium, the price of milk exceeds average total cost, so the firms are making positive profit.

FIGURE 6.14**An Increase in Demand in the Short Run and Long Run**

The market starts in a long-run equilibrium, shown as point A in panel (a). In this equilibrium, each firm makes zero profit, and the price equals the minimum average total cost. Panel (b) shows what happens in the short run when demand rises from D_1 to D_2 . The equilibrium goes from point A to point B, price rises from P_1 to P_2 , and the quantity sold in the market rises from Q_1 to Q_2 . Because price now exceeds average total cost, firms make profits, which over time encourage new firms to enter the market. This entry shifts the short-run supply curve to the right from S_1 to S_2 , as shown in panel (c). In the new long-run equilibrium, point C, price has returned to P_1 , but the quantity sold has increased to Q_3 . Profits are again zero, price is back to the minimum of average total cost, but the market has more firms to satisfy the greater demand.



Over time, the profit in this market encourages new firms to enter. Some farmers may switch to milk from other farm products, for example. As the number of firms grows, the short-run supply curve shifts to the right from S_1 to S_2 , as in panel (c), and this shift causes the price of milk to fall. Eventually, the price is driven back down to the minimum of average total cost, profits are zero and firms stop entering. Thus, the market reaches a new long-run equilibrium, point C. The price of milk has returned to P_1 , but the quantity produced has risen to Q_3 . Each firm is again producing at its efficient scale, but, because more firms are in the dairy business, the quantity of milk produced and sold is higher.

Why the Long-Run Supply Curve Might Slope Upwards

So far we have seen that entry and exit can cause the long-run market supply curve to be horizontal. The essence of our analysis is that there are a large number of potential entrants, each of which faces the same costs. As a result, the long-run market supply curve is horizontal at the minimum of average total cost. When the demand for the good increases, the long-run result is an increase in the number of firms and in the total quantity supplied, without any change in the price.

There are, however, two reasons that the long-run market supply curve might slope upward.

Some Resources Used in Production May Be Available Only in Limited Quantities For example, consider the market for farm products. Anyone can choose to buy land and start a farm, but the quantity and quality of land is limited. As more people become farmers, the price of farm land is bid up, which raises the costs of all farmers in the market. Thus, an increase in demand for farm products cannot induce an increase in quantity supplied without also inducing a rise in farmers' costs, which in turn means a rise in price. The result is a long-run market supply curve that is upwards sloping, even with free entry into farming.

Firms May Have Different Costs Consider the market for painters. Anyone can enter the market for painting services, but not everyone has the same costs. Costs vary in part because some people work faster than others, use different materials and equipment and because some people have better alternative uses of their time than others. For any given price, those with lower costs are more likely to enter than those with higher costs. To increase the quantity of painting services supplied, additional entrants must be encouraged to enter the market. Because these new entrants have higher costs, the price must rise to make entry profitable for them. Thus, the market supply curve for painting services slopes upwards even with free entry into the market.

Notice that if firms have different costs, some firms earn profit even in the long run. In this case, the price in the market reflects the average total cost of the *marginal firm* – the firm that would exit the market if the price were any lower. This firm earns zero profit, but firms with lower costs earn positive profit. Entry does not eliminate this profit because would-be entrants have higher costs than firms already in the market. Higher-cost firms will enter only if the price rises, making the market profitable for them.

Thus, for these two reasons, the long-run supply curve in a market may be upwards sloping rather than horizontal, indicating that a higher price is necessary to induce a larger quantity supplied. Nevertheless, the basic lesson about entry and exit remains true. Because firms can enter and exit more easily in the long run than in the short run, the long-run supply curve is typically more elastic than the short-run supply curve.

SELF TEST In the long run with free entry and exit, is the price in a market equal to marginal cost, average total cost, both, or neither? Explain with a diagram.

CONCLUSION: BEHIND THE SUPPLY CURVE

We have been discussing the behaviour of competitive profit-maximizing firms. You may recall from Chapter 1 that one of the *Ten Principles of Economics* is that rational people think at the margin. This chapter has applied this idea to the competitive firm. Marginal analysis has given us a theory of the supply curve in a competitive market and, as a result, a deeper understanding of market outcomes.

We have learned that when you buy a good from a firm in a competitive market, you can be assured that the price you pay is close to the cost of producing that good. In particular, if firms are competitive and

profit-maximizing, the price of a good equals the marginal cost of making that good. In addition, if firms can freely enter and exit the market, the price also equals the lowest possible average total cost of production.

In later chapters we will examine the behaviour of firms with market power. Marginal analysis will again be useful in analysing these firms, but it will have quite different implications.

IN THE NEWS

Marginal Cost and Prices

Our look at the supply curve in this chapter has been based on a model of a perfectly competitive firm in which there are a number of assumptions. One of the conclusions of the analysis is that under perfectly competitive conditions a firm charges a price which equals marginal cost. Does this mean that if the marginal cost is virtually zero that price should also be virtually zero?

The Marginal Cost of Text Messaging

How much do you pay to send a text message? The amount varies depending on the type of contract you are on and the service provider you are with. Given the millions of text messages sent have you ever

With 1/1000th of a cent being the average cost of transferring a text message, are network providers exploiting customers by charging them an average price of €0.10–€0.15 per text message?



considered what the cost of transferring a text message to a service provider is? Surely it is not very much. If it is a small amount then are service providers making large profits from consumers who send text messages?

One estimate has put the cost of transferring a text message at around 1/1000th of a cent. If this is the case, then with the average price to a consumer of sending a text message being something around €0.10–€0.15 then firms must be making considerable profits – are network providers exploiting customers and making abnormal profits?

What we are looking at here is the marginal cost to a service provider of transferring a text message for a customer. Whether the marginal cost is 1/1000th of a cent or, as other people have suggested, virtually zero, is debatable. The network handling the transfer will be likely to experience some energy costs in making the transfer and there will be some way in which the message is recorded so that the customer pays the correct amount on their bill. There will be some

additional cost to the network provider but it is likely to be tiny. So, if we agree that the marginal cost is almost zero why isn't the price to the consumer almost zero as well? Given the price consumers pay for text messages does this mean consumers are being exploited?

The answer is – possibly they are but what does need to be considered are the fixed costs associated with transferring text messages. In the UK, network providers handed over £20 billion to buy the licence to run 3G services. In addition, these firms have had considerable capital investment to run these telecommunications services even if text messaging is just one of the services that rides on the back of the network infrastructure that telecoms firms have had to pay to set up.

The difference in the price paid by the consumer and the cost to the network of sending a text message, is therefore, likely to be a contribution to these fixed costs. Firms will have to find a way of assessing the size of this contribution and setting the price to consumers.

Questions

- 1 Why is the marginal cost of sending a text message likely to be extremely low?

- 2** If network providers charge consumers €0.15 to send a text message but the marginal cost of sending a text is €0.001, what is the marginal revenue of every text sent?
- 3** If network providers are making short run abnormal profits from text messaging services, what would you expect to happen in the long run in this industry?
- 4** What factors might prevent the long-run position in this industry being different to that in a perfectly competitive market as described in this chapter?
- 5** The article suggests that the fixed costs in this industry are very high. Does this affect the profit maximizing output for network providers. Explain.

SUMMARY

- When analysing a firm's behaviour, it is important to include all the opportunity costs of production. Some, such as the wages a firm pays its workers, are explicit. Others, such as the wages the firm owner gives up by working in the firm rather than taking another job, are implicit.
- A firm's costs reflect its production process. A typical firm's production function gets flatter as the quantity of an input increases, displaying the property of diminishing marginal product. As a result, a firm's total cost curve gets steeper as the quantity produced rises.
- A firm's total costs can be divided between fixed costs and variable costs. Fixed costs are costs that are not determined by the quantity of output produced. Variable costs are costs that directly relate to the amount produced and so change when the firm alters the quantity of output produced.
- Average total cost is total cost divided by the quantity of output. Marginal cost is the amount by which total cost changes if output increases (or decreases) by 1 unit.
- For a typical firm, marginal cost rises with the quantity of output. Average total cost first falls as output increases and then rises as output increases further. The marginal cost curve always crosses the average total cost curve at the minimum of average total cost.
- Many costs are fixed in the short run but variable in the long run. As a result, when the firm changes its level of production, average total cost may rise more in the short run than in the long run.
- Because a competitive firm is a price taker, its revenue is proportional to the amount of output it produces. The price of the good equals both the firm's average revenue and its marginal revenue.
- One goal of firms is to maximize profit, which equals total revenue minus total cost.
- To maximize profit, a firm chooses a quantity of output such that marginal revenue equals marginal cost. Because marginal revenue for a competitive firm equals the market price, the firm chooses quantity so that price equals marginal cost. Thus, the firm's marginal cost curve is its supply curve.
- In the short run when a firm cannot recover its fixed costs, the firm will choose to shut down temporarily if the price of the good is less than average variable cost. In the long run when the firm can recover both fixed and variable costs, it will choose to exit if the price is less than average total cost.
- In a market with free entry and exit, profits are driven to zero in the long run. In this long-run equilibrium, all firms produce at the efficient scale, price equals the minimum of average total cost, and the number of firms adjusts to satisfy the quantity demanded at this price.
- Changes in demand have different effects over different time horizons. In the short run, an increase in demand raises prices and leads to profits, and a decrease in demand lowers prices and leads to losses. But if firms can freely enter and exit the market, then in the long run the number of firms adjusts to drive the market back to the zero-profit equilibrium.

QUESTIONS FOR REVIEW

- 1 What is marginal product, and what does it mean if it is diminishing?
- 2 Draw a production function that exhibits diminishing marginal product of labour. Draw the associated total cost curve. (In both cases, be sure to label the axes.) Explain the shapes of the two curves you have drawn.
- 3 Give an example of an opportunity cost that an accountant might not count as a cost. Why would the accountant ignore this cost?

- 4 Draw the cost curves for a typical firm. For a given price, explain how the firm chooses the level of output that maximizes profit.
- 5 Draw the marginal cost and average total cost curves for a typical firm. Explain why the curves have the shapes that they do and why they cross where they do.
- 6 Under what conditions will a firm shut down temporarily? Explain.
- 7 Under what conditions will a firm exit a market? Explain.
- 8 Does a firm's price equal marginal cost and the minimum of average total cost in the short run, in the long run, or both? Explain.
- 9 Explain the difference between increasing, constant and decreasing returns to scale.
- 10 Are market supply curves typically more elastic in the short run or in the long run? Explain.

PROBLEMS AND APPLICATIONS

- 1 Manton Bakery is a company that bakes bread. Here is the relationship between the number of workers at the bakery and Manton's output in a given day:

| Workers | Output of loaves | Marginal product | Average total cost | Marginal cost |
|---------|------------------|------------------|--------------------|---------------|
| 0 | 0 | | | |
| 1 | 20 | | | |
| 2 | 50 | | | |
| 3 | 90 | | | |
| 4 | 120 | | | |
| 5 | 140 | | | |
| 6 | 150 | | | |
| 7 | 155 | | | |

- a. Fill in the column of marginal product. What pattern do you see? How might you explain it?
 - b. A skilled baker costs €100 a day, and the firm has fixed costs of €200. Use this information to fill in the column for total cost.
 - c. Fill in the column for average total cost. (Recall that $ATC = TC/Q$.) What pattern do you see?
 - d. Now fill in the column for marginal cost. (Recall that $MC = \Delta TC/\Delta Q$.) What pattern do you see?
 - e. Compare the column for marginal product and the column for marginal cost. Explain the relationship.
 - f. Compare the column for average total cost and the column for marginal cost. Explain the relationship.
- 2 Your aunt announces that she is thinking about opening a restaurant. She estimates that it would cost €500,000 per year to rent the premises, buy a licence to serve alcohol and to buy in food. In addition, she would have to leave her €50,000 per year job as an accountant.
 - a. Define opportunity cost.
 - b. What is your aunt's opportunity cost of running the restaurant for a year? If your aunt thought she could sell €510,000 worth of food in a year, should she open the restaurant? Explain.
 - 3 Your cousin Vinnie owns a painting company with fixed costs of €200 and the following schedule for variable costs:

| Quantity of houses painted per month | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------------------------------------|-----|-----|-----|-----|------|------|------|
| Variable costs | €10 | €20 | €40 | €80 | €160 | €320 | €640 |

Calculate average fixed cost, average variable cost and average total cost for each quantity. What is the efficient scale of Vinnie's company?

- 4 What are the characteristics of a competitive market? Which of the following drinks do you think is best described by these characteristics? Why aren't the others?
 - a. tap water
 - b. bottled water
 - c. cola
 - d. beer.
- 5 Your flatmate's long hours in the chemistry lab finally paid off – she discovered a secret formula that lets people do an hour's worth of studying in 5 minutes. So far, she's sold 200 doses, and faces the following average total cost schedule:

| Q | Average total cost (€) |
|-----|------------------------|
| 199 | 199 |
| 200 | 200 |
| 201 | 201 |

If a new customer offers to pay your flatmate €300 for one dose, should she make one more? Explain.

- 6 You go out to the best restaurant in town and order a steak tartar dinner for €40. After eating half of the steak tartar, you realize that you are quite full. Your date wants you to finish your dinner, because you can't take it home and because 'you've already paid for it'. What should you do? Relate your answer to the material in this chapter.
- 7 Alejandro's lawn-mowing service is a profit-maximizing, competitive firm. Alejandro mows lawns for €27 each. His total cost each day is €280, of which €30 is a fixed cost. He mows 10 lawns a day. What can you say about Alejandro's short-run decision regarding shutdown and his long-run decision regarding exit?

8 Consider total cost and total revenue given in the table below:

| Quantity | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------------|----|----|-----|-----|-----|-----|-----|-----|
| Total cost | €8 | 19 | 110 | 111 | 113 | 119 | 127 | 137 |
| Total revenue | €0 | 8 | 16 | 24 | 32 | 40 | 48 | 56 |

- Calculate profit for each quantity. How much should the firm produce to maximize profit?
- Calculate marginal revenue and marginal cost for each quantity. Graph them. (Hint: put the points between whole numbers. For example, the marginal cost between 2 and 3 should be graphed at $2\frac{1}{2}$.) At what quantity do these curves cross? How does this relate to your answer to part (a)?
- Can you tell whether this firm is in a competitive industry? If so, can you tell whether the industry is in long-run equilibrium?

9 'High prices traditionally cause expansion in an industry, eventually bringing an end to high prices and manufacturers' prosperity.' Explain, using appropriate diagrams.

10 Suppose the book-printing industry is competitive and begins in long-run equilibrium.

- Draw a diagram describing the typical firm in the industry.
- Hi-Tech Printing Company invents a new process that sharply reduces the cost of printing books. What happens to Hi-Tech's profits and the price of books in the short run when Hi-Tech's patent prevents other firms from using the new technology?
- What happens in the long run when the patent expires and other firms are free to use the technology?

PART 3

MARKETS, EFFICIENCY AND WELFARE

7

CONSUMERS, PRODUCERS AND THE EFFICIENCY OF MARKETS

In Chapters 3 and 4 we looked at the working of the free market and concluded that given the assumptions made, free markets address the three key questions all economies have to answer – what is to be produced, how are the goods and services produced and who gets what is produced? In Chapters 5 and 6 we looked in more detail at the demand and supply side of the free market and how the demand and supply curve are derived.

Having described the way markets allocate scarce resources we now need to address the question of whether these market allocations are desirable. In other words, our analysis has been *positive* (what is) rather than *normative* (what should be). We know that the price of a good adjusts to ensure that the quantity of a good supplied equals the quantity demanded. But, at this equilibrium, is the quantity of the good produced and consumed too small, too large or just right?

In this chapter we take up the topic of **welfare economics**, the study of how the allocation of resources affects economic **well-being**. Economists use the term well-being a good deal and have taken steps to define the term. A UK Treasury Economic Working Paper published in 2008 (Lepper, L. and McAndrew, S. (2008) *Developments in the Economics of Well-being*. Treasury Working Paper number 4) highlighted two main definitions of economic well-being – subjective and objective well-being. Subjective well-being refers to the way in which people evaluate their own happiness. This includes how they feel about work, leisure and their response to the events which occur in their lives. Objective well-being refers to measures of the quality of life and uses indicators developed by researchers such as educational attainment, measures of the standard of living, life expectancy and so on. Welfare economics uses some of the microeconomic techniques we have already looked at to estimate **allocative efficiency** – a measure of the utility

welfare economics the study of how the allocation of resources affects economic well-being

well-being happiness or satisfaction with life as reported by individuals

allocative efficiency a resource allocation where the value of the output by sellers matches the value placed on that output by buyers

(satisfaction) derived from the allocation of resources. We have seen how buyers place a value on consumption in Chapter 5. Allocative efficiency occurs when the value of the output that firms produce (the benefits to sellers) matches the value placed on that output by consumers (the benefit) to buyers.

We begin by examining the benefits that buyers and sellers receive from taking part in a market. We then examine how society can make these benefits as large as possible. This analysis leads to a profound conclusion: the equilibrium of supply and demand in a market maximizes the total benefits received by buyers and sellers.

One of the *Ten Principles of Economics* is that markets are usually a good way to organize economic activity. The study of welfare economics explains this principle more fully. It also answers our question about the right price of a good: the price that balances the supply and demand for any good is, in a particular sense, the best one because it maximizes the total welfare of consumers and producers.

CONSUMER SURPLUS

We begin our study of welfare economics by looking at the benefits buyers receive from participating in a market.

Willingness to Pay

Imagine that you own an extremely rare signed vintage electric guitar which you decide to sell. One way to do so is to hold an auction.

Four guitar collectors show up for your auction: Liam, Paul, Noel and Tony. Each of them would like to own the guitar, but there is a limit to the amount that each is willing to pay for it. Table 7.1 shows the maximum price that each of the four possible buyers would pay. Each buyer's maximum is called his **willingness to pay**, and it measures how much that buyer values the good. Each buyer has his own value assigned to the guitar, which is expressed as the price he is willing to pay to own it. Each will have some upper limit above which they will not be prepared to pay (possibly because they don't feel the guitar is worth it above that upper limit or because they know they cannot afford to pay any more). If the price were below this upper limit then each would be eager to buy the guitar.

willingness to pay the maximum amount that a buyer will pay for a good

TABLE 7.1 Four Possible Buyers' Willingness to Pay

| Buyer | Willingness to pay (€) |
|-------|------------------------|
| Liam | 1,000 |
| Paul | 800 |
| Noel | 700 |
| Tony | 500 |

To sell your guitar, you begin the bidding at a low price, say €100. Because all four buyers are willing to pay much more, the price rises quickly. The bidding stops when Liam bids €801. At this point, Paul, Noel and Tony have dropped out of the bidding because they are unwilling to bid any more than €800. Liam pays you €801 and gets the guitar. Note that the guitar has gone to the buyer who values it most highly.

What benefit does Liam receive from buying the guitar? In a sense, Liam says he has 'found a real bargain': he is willing to pay €1,000 for the guitar but pays only €801 for it. Liam valued the benefits from owning the guitar more highly than the money he has had to give up to own it. One way to express the

value of these benefits is in monetary terms. We say that Liam receives *consumer surplus* of €199. **Consumer surplus** is the amount a buyer is willing to pay for a good minus the amount the buyer actually pays for it. We refer to ‘getting a bargain’ regularly in everyday language. In economics, a bargain means paying much less for something than we expected or anticipated and as a result we get a greater degree of consumer surplus than we expected.

consumer surplus a buyer’s willingness to pay minus the amount the buyer actually pays

Consumer surplus measures the benefit to buyers of participating in a market. In this example, Liam receives a €199 benefit from participating in the auction because he pays only €801 for a good he values at €1,000. Paul, Noel and Tony get no consumer surplus from participating in the auction because they left without the guitar and without paying anything.

Now consider a somewhat different example. Suppose that you had two identical guitars to sell. Again, you auction them off to the four possible buyers. To keep things simple, we assume that both guitars are to be sold for the same price and that no buyer is interested in buying more than one guitar. Therefore, the price rises until two buyers are left.

In this case, the bidding stops when Liam and Paul bid €701. At this price, Liam and Paul are each happy to buy a guitar and Noel and Tony are not willing to bid any higher. Liam and Paul each receive consumer surplus equal to his willingness to pay minus the price. Liam’s consumer surplus is €299 and Paul’s is €99. Liam’s consumer surplus is higher now than it was previously, because he gets the same guitar but pays less for it. The total consumer surplus in the market is €398.

Using the Demand Curve to Measure Consumer Surplus

Consumer surplus is closely related to the demand curve for a product. To see how they are related, let’s continue our example and consider the demand curve for guitars.

We begin by using the willingness to pay of the four possible buyers to find the demand schedule for the album. The table in Figure 7.1 shows the demand schedule that corresponds to Table 7.1. If the price is above €1,000, the quantity demanded in the market is 0, because no buyer is willing to pay that much. If the price is between €801 and €1,000, the quantity demanded is 1, because only Liam is willing to pay such a high price. If the price is between €701 and €801, the quantity demanded is 2, because both Liam and Paul are willing to pay the price. We can continue this analysis for other prices as well. In this way, the demand schedule is derived from the willingness to pay of the four possible buyers.

The graph in Figure 7.1 shows the demand curve that corresponds to this demand schedule. Note the relationship between the height of the demand curve and the buyers’ willingness to pay. At any quantity, the price given by the demand curve shows the willingness to pay of the *marginal buyer*, the buyer who would leave the market first if the price were any higher. At a quantity of 4 guitars, for instance, the demand curve has a height of €500, the price that Tony (the marginal buyer) is willing to pay for a guitar. At a quantity of 3 guitars, the demand curve has a height of €700, the price that Noel (who is now the marginal buyer) is willing to pay.

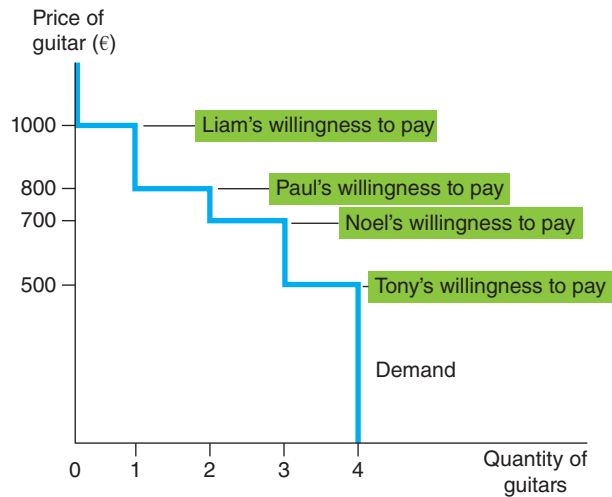
Because the demand curve reflects buyers’ willingness to pay, we can also use it to measure consumer surplus. Figure 7.2 uses the demand curve to compute consumer surplus in our example. In panel (a), the price is €801 and the quantity demanded is 1. Note that the area above the price and below the demand curve equals €199 ($€1,000 - €801 \times 1$). This amount is exactly the consumer surplus we computed earlier when only 1 guitar is sold.

Panel (b) of Figure 7.2 shows consumer surplus when the price is €701. In this case, the area above the price and below the demand curve equals the total area of the two rectangles: Liam’s consumer surplus at this price is €299 and Paul’s is €99. This area equals a total of €398. Once again, this amount is the consumer surplus we computed earlier.

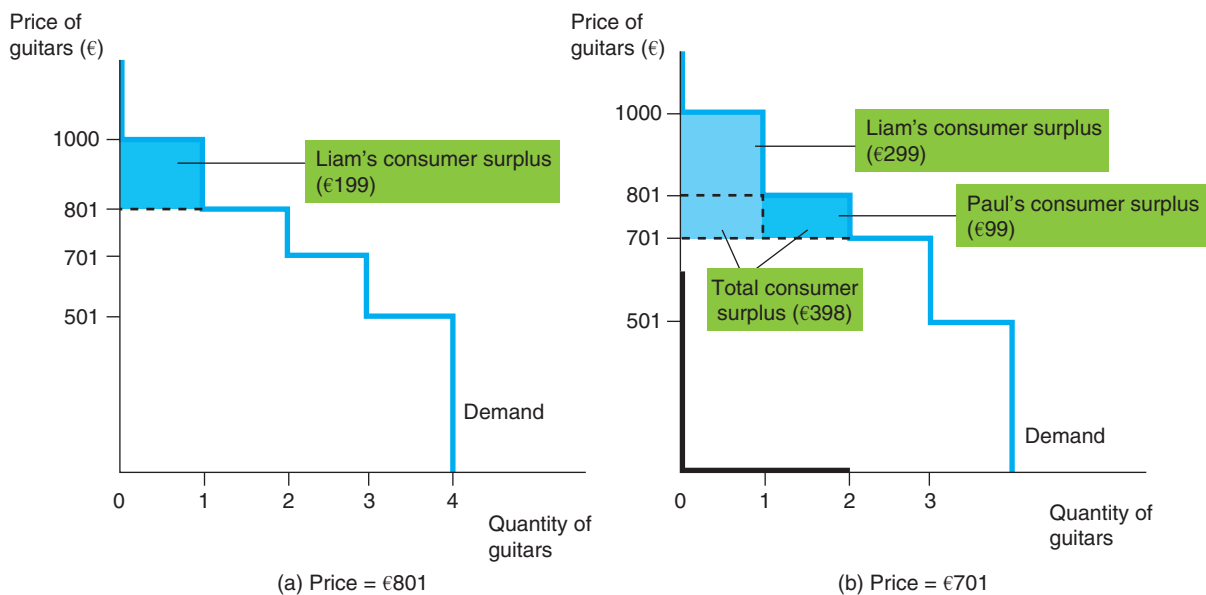
FIGURE 7.1**The Demand Schedule and the Demand Curve**

The table shows the demand schedule for the buyers in Table 7.1. The graph shows the corresponding demand curve. Note that the height of the demand curve reflects buyers' willingness to pay.

| Price | Buyers | Quantity demanded |
|------------------|------------------------|-------------------|
| More than €1,000 | None | 0 |
| €801 to €1,000 | Liam | 1 |
| €701 to €801 | Liam, Paul | 2 |
| €501 to €701 | Liam, Paul, Noel | 3 |
| €500 or less | Liam, Paul, Noel, Tony | 4 |

**FIGURE 7.2****Measuring Consumer Surplus with the Demand Curve**

In panel (a) the price of the good is €801 and the consumer surplus is €199. In panel (b) the price of the good is €701 and the consumer surplus is €398.



The lesson from this example holds for all demand curves: the area below the demand curve and above the price measures the consumer surplus in a market. The reason is that the height of the demand curve multiplied by the quantity measures the value buyers place on the good, as represented by their willingness to pay for it. The difference between this willingness to pay and the market price is each buyer's consumer surplus. Thus, the total area below the demand curve and above the price is the sum of the consumer surplus of all buyers in the market for a good or service.

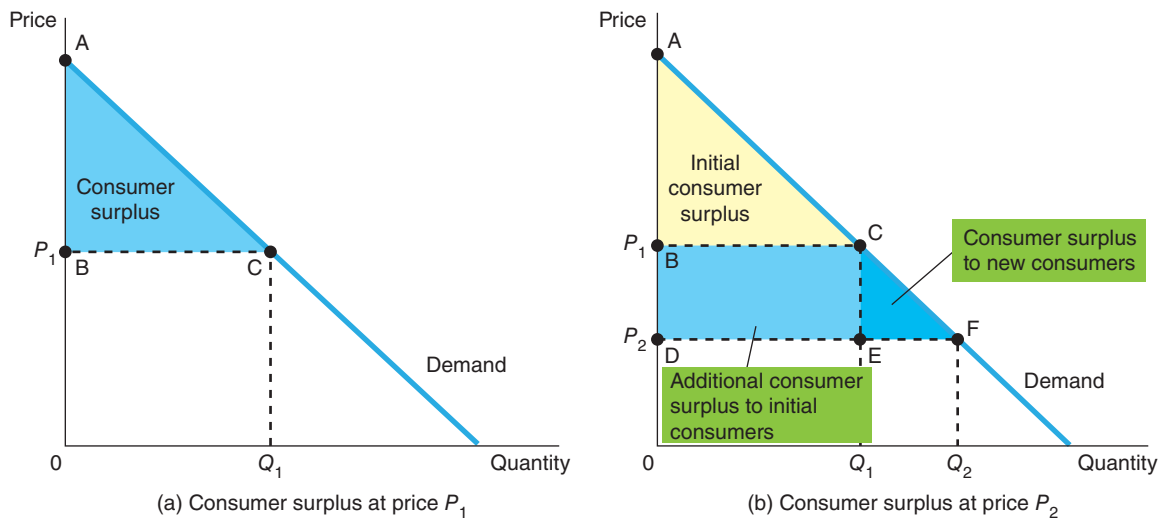
How a Lower Price Raises Consumer Surplus

Because buyers always want to pay less for the goods they buy, a lower price makes buyers of a good better off and improves their well-being. Figure 7.3 shows a typical downwards sloping demand curve. Although this demand curve appears somewhat different in shape from the step-like demand curves in our previous two figures, the ideas we have just developed apply nevertheless: consumer surplus is the area above the price and below the demand curve. In panel (a), consumer surplus at a price of P_1 is the area of triangle ABC.

FIGURE 7.3

How the Price Affects Consumer Surplus

In panel (a) the price is P_1 , the quantity demanded is Q_1 , and consumer surplus equals the area of the triangle ABC. When the price falls from P_1 to P_2 , as in panel (b), the quantity demanded rises from Q_1 to Q_2 , and the consumer surplus rises to the area of the triangle ADF. The increase in consumer surplus (area BCFD) occurs in part because existing consumers now pay less (area BCED) and in part because new consumers enter the market at the lower price (area CEF).



Now suppose that the price falls from P_1 to P_2 , as shown in panel (b). The consumer surplus now equals area ADF. The increase in consumer surplus attributable to the lower price is the area BCFD.

This increase in consumer surplus is composed of two parts. First, those buyers who were already buying Q_1 of the good at the higher price P_1 are better off because they now pay less. The increase in consumer surplus of existing buyers is the reduction in the amount they pay; it equals the area of the rectangle BCED. Second, some new buyers enter the market because they are now willing to buy the good at the lower price. As a result, the quantity demanded in the market increases from Q_1 to Q_2 . The consumer surplus these newcomers receive is the area of the triangle CEF.

What Does Consumer Surplus Measure?

Our goal in developing the concept of consumer surplus is to make normative judgements about the desirability of market outcomes. Imagine that you are a policymaker trying to design a good economic system. Consumer surplus would be important to consider as it measures the net economic benefit in terms of surplus value that buyers receive from a good *as the buyers themselves perceive it*. The demand curve is a representation of the value of the economic benefit consumers get from consumption. This value is related to the amount of the consumer's scarce income they are willing to give up to purchase the good represented by the price they have to pay to acquire the good. Every consumer

(mostly unconsciously) weighs up the value to them of buying a good. Psychologists have shown that there are lots of different things going on when we make such choices apart from simply a rational weighing up of the costs and benefits as we saw in our discussion of heuristics in Chapter 5. As a consumer yourself, you will almost certainly be able to bring to mind instances where you have agonized over whether to buy something and if you were asked at that moment to describe your thinking you would no doubt be weighing up a variety of factors. If you are agonizing then you are operating right at this marginal value – the maximum amount you are prepared to pay. For some reason, if the price you are being asked to pay is slightly higher you decide not to buy – what you are being asked to give up is not offset by the value of the benefit you perceive you will get from purchasing the good.

You might also recall times when you have seen a good and snapped it up – you think to yourself you have a bargain. You now have the tools to understand why you experience that feeling of getting a bargain – it is because of the amount of consumer surplus you have gained from the purchase. Thus, consumer surplus is a good measure of economic well-being if policymakers want to respect the preferences of buyers.

Conceiving of Price as a Bargaining Model Our discussion of markets so far has noted that price acts as a signal to buyers and sellers. The actual purchasing decision by a consumer can be seen from the perspective of a bargaining model. Suppliers are offering goods to consumers at different prices and consumers have to make decisions about whether the prices they are offered represent a net economic benefit to them. There is an interaction between suppliers and consumers, therefore, which can be seen as a bargaining process – an agreed outcome between two interested and competing economic agents.

Think of the times when you have trawled through a price comparison website or been around almost every shop in a mall only to return to the first item you saw in the first shop you went into and bought that item. In these cases consumers are responding to the prices being offered by suppliers and making decisions based on the competing prices available. Suppliers respond to the decisions made by consumers – if too few people buy their product then they will be forced to take action to improve the product offering. If consumers buy the product in sufficient numbers to make it worthwhile for producers, then this implies that the supplier has some understanding of the net benefit to consumers and can continue to work on finding ways to maximize this benefit at prices consumers are willing to pay.

Is Consumer Surplus Always a Good Measure of Economic Well-Being?

In some circumstances, policymakers might choose not to care about consumer surplus because they do not respect the preferences that drive buyer behaviour. For example, drug addicts are willing to pay a high price for heroin. Yet we would not say that addicts get a large benefit from being able to buy heroin at a low price (even though addicts might say they do). From the standpoint of society, willingness to pay, in this instance, is not a good measure of the buyers' benefit, and consumer surplus is not a good measure of economic well-being, because addicts are not looking after their own best interests.

The use of the word 'good' to describe products has not just arisen as a result of chance. A product described as a 'good' implies that consumption of them confers positive benefits on consumers. Products like non-medicinal drugs, tobacco and alcohol might be better described as 'bads' rather than 'goods' because they confer negative benefits to the consumer such as a deterioration in long-term physical and mental health even though many consumers of these goods would claim that they enjoy consuming them.

In most markets, however, consumer surplus does reflect economic well-being. The underlying assumption to this is that we are presuming that buyers are rational when they make decisions and that their preferences should be respected. In this case, consumers are the best judges of how much benefit they receive from the goods they buy. As we have seen, this assumption is open to some debate. In addition it must be noted that an assumption made in the analysis so far is that one person's value of an extra unit of a euro is the same as someone else's. For example, if the price of a guitar in an auction rose by €1 from €750 to €751, do Paul and Liam place the same value on that extra euro? In this analysis we are assuming they are but in reality this may not be the case.

As with many things in economics, we are attempting to introduce some basic principles of the subject but, as you progress through your studies to intermediate and advanced levels, you will find that the simplified assumptions that are made can (and should) be challenged and new, more sophisticated understandings begin to emerge.

SELF TEST Think about an occasion when you have used an auction website such as eBay. How does the principle behind eBay exemplify the concept of consumer surplus?

CASE STUDY Imported Car Policies and Consumer Surplus in Pakistan

According to the World Bank, in 2011 Pakistan had a population of about 177 million people and around a quarter of the population lived below the poverty line. One of the problems its citizens face is transport – public transport and mass-transit infrastructure are limited, which means that large numbers of people are dependent on cars for transport. Pakistan has a local car assembly industry dominated by three major companies: Indus Motor Company Limited (a division of Toyota), Pak-Suzuki Motor Company Limited and Honda Atlas Cars Limited. These companies had been complaining that their business was being negatively affected by imported cars and so the government imposed new rules for imported cars in 2012.

The new rules place a limit on the age of cars which can be imported into the country. Previously, importers were allowed to bring five-year old cars into Pakistan but this has been reduced to three-year old cars. For the people of Pakistan, the choices they face now change. For many people, a second-hand car is the only option to help their transport problems but three-year old imported cars will be more expensive than the five-year old cars which they had access to prior to the new rules. They can, of course, switch to buying the locally produced cars but many complain that the quality of these vehicles is not as high as imported cars and, in addition, the new rules would reduce competition and allow the three dominant firms to increase prices. In one class of vehicle, the 650cc class, consumers had a choice of 18 different models when five-year old imported cars were available but the new rules would reduce the choice to just one local model in that particular price range.

Either way, the effect of the new rules will be to reduce the consumer surplus of some Pakistan citizens by having a disproportionate effect on the poor. Those in society who are relatively wealthy will be able to afford to buy the higher priced imported vehicles but the higher prices overall will reduce overall welfare because of the high proportion of poor and low income families in the country.



The car in Pakistan is an important source of transport – government policy designed to protect local car manufacturers might adversely affect consumers, lowering consumer surplus and welfare.

PRODUCER SURPLUS

We now turn to the other side of the market and consider the benefits sellers receive from participating in a market. As you will see, our analysis of sellers' welfare is similar to our analysis of buyers' welfare.

Cost and the Willingness to Sell

Imagine now that you own a house and you need to get it painted externally. You turn to four sellers of house painting services: Millie, Georgia, Julie and Nana. Each painter is willing to do the work for you if the price is right. You decide to take bids from the four painters and auction off the job to the painter who will do the work for the lowest price (assuming the quality each painter provides is the same).

Each painter is willing to take the job if the price she would receive exceeds her **cost** of doing the work. Here the term cost should be interpreted as the painter's opportunity cost: it includes the painter's out-of-pocket expenses (for paint, brushes and so on) as well as the value that the painter places on her own time. Table 7.2 shows each painter's cost. Because a painter's cost is the lowest price she would accept for her work, cost is a measure of her willingness to sell her services. Each painter would be eager to sell her services at a price greater than her cost, would refuse to sell her services at a price less than her cost, and would be indifferent about selling her services at a price exactly equal to her cost.

cost the value of everything a seller must give up to produce a good

TABLE 7.2 The Costs of Four Possible Sellers

| Seller | Cost (€) |
|---------|----------|
| Millie | 900 |
| Julie | 800 |
| Georgia | 600 |
| Nana | 500 |

When you take bids from the painters, the price might start off high, but it quickly falls as the painters compete for the job. Once Nana has bid slightly less than €600 (€599.99), she is the sole remaining bidder. Nana is happy to do the job for this price, because her cost is only €500. Millie, Georgia and Julie are unwilling to do the job for less than €600. Note that the job goes to the painter who can do the work at the lowest cost.

What benefit does Nana receive from getting the job? Because she is willing to do the work for €500 but gets €599.99 for doing it, we say that she receives *producer surplus* of €99.99. **Producer surplus** is the amount a seller is paid minus the cost of production. Producer surplus measures the benefit to sellers of participating in a market.

producer surplus the amount a seller is paid for a good minus the seller's cost

Now suppose that you have two houses that need painting. Again, you auction off the jobs to the four painters. To keep things simple, let's assume that no painter is able to paint both houses and that you will pay the same amount to paint each house. Therefore, the price falls until two painters are left.

In this case, the bidding stops when Georgia and Nana each offer to do the job for a price slightly less than €800 (€799.99). At this price, Georgia and Nana are willing to do the work, and Millie and Julie are not willing to bid a lower price. At a price of €799.99, Nana receives producer surplus of €299.99, and Georgia receives producer surplus of €199.99. The total producer surplus in the market is €499.98.

Using the Supply Curve to Measure Producer Surplus

Just as consumer surplus is closely related to the demand curve, producer surplus is closely related to the supply curve. To see how, let's continue our example.

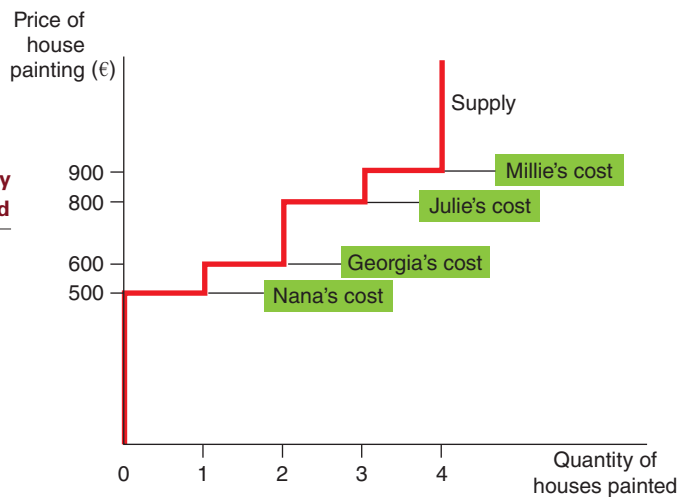
We begin by using the costs of the four painters to find the supply schedule for painting services. The table in Figure 7.4 shows the supply schedule that corresponds to the costs in Table 7.2. If the price is below €500, none of the four painters is willing to do the job, so the quantity supplied is zero. If the price is between €500 and €599, only Nana is willing to do the job, so the quantity supplied is 1. If the price is between €600 and €799, Nana and Georgia are willing to do the job, so the quantity supplied is 2, and so on. Thus, the supply schedule is derived from the costs of the four painters.

FIGURE 7.4

The Supply Schedule and the Supply Curve

The table shows the supply schedule for the sellers in Table 7.2. The graph shows the corresponding supply curve. Note that the height of the supply curve reflects sellers' costs.

| Price | Sellers | Quantity supplied |
|----------------|------------------------------|-------------------|
| €901 or more | Millie, Julie, Georgia, Nana | 4 |
| €801 to €900 | Julie, Georgia, Nana | 3 |
| €601 to €800 | Georgia, Nana | 2 |
| €500 to €600 | Nana | 1 |
| Less than €500 | None | 0 |



The graph in Figure 7.4 shows the supply curve that corresponds to this supply schedule. Note that the height of the supply curve is related to the sellers' costs. At any quantity, the price given by the supply curve shows the cost of the *marginal seller*, the seller who would leave the market first if the price were any lower. At a quantity of 4 houses, for instance, the supply curve has a height of €900, the cost that Millie (the marginal seller) incurs to provide her painting services. At a quantity of 3 houses, the supply curve has a height of €800, the cost that Julie (who is now the marginal seller) incurs.

Because the supply curve reflects sellers' costs, we can use it to measure producer surplus. Figure 7.5 uses the supply curve to compute producer surplus in our example. In panel (a) we assume that the price is €599.99. In this case, the quantity supplied is 1. Note that the area below the price and above the supply curve equals €99.99. This is Nana's producer surplus.

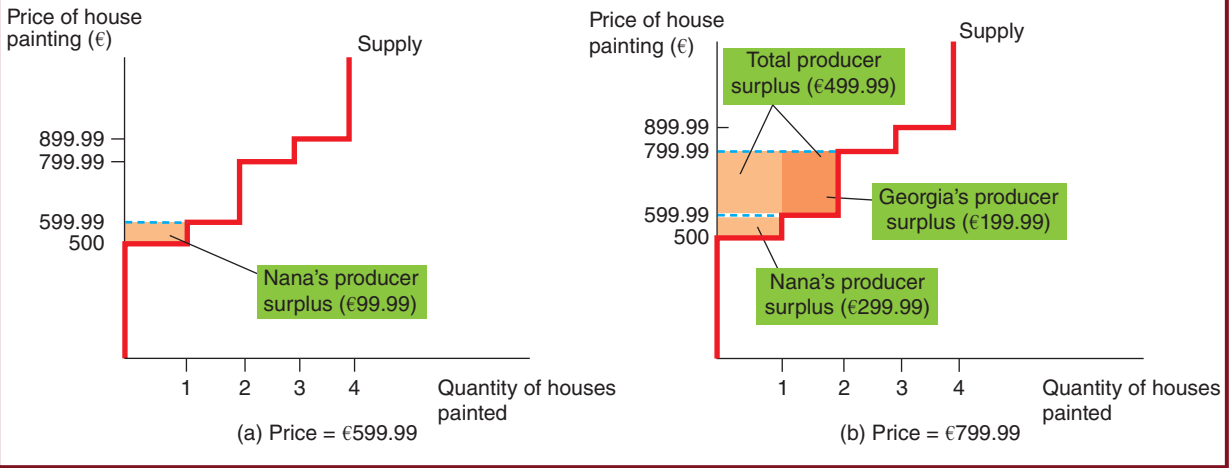
Panel (b) of Figure 7.5 shows producer surplus at a price of €799.99. In this case, the area below the price and above the supply curve equals the total area of the two rectangles. This area equals €499.99, the producer surplus we computed earlier for Georgia and Nana when two houses needed painting.

The lesson from this example applies to all supply curves: the area below the price and above the supply curve measures the producer surplus in a market. The logic is straightforward: the height of the supply curve measures sellers' costs, and the difference between the price and the cost of production is each seller's producer surplus. When multiplied by the quantity, the total area is the sum of the producer surplus of all sellers.

FIGURE 7.5

Measuring Producer Surplus with the Supply Curve

In panel (a) the price of the good is €599.99 and the producer surplus is €99.99. In panel (b) the price of the good is €799.99 and the producer surplus is €499.99.



How a Higher Price Raises Producer Surplus

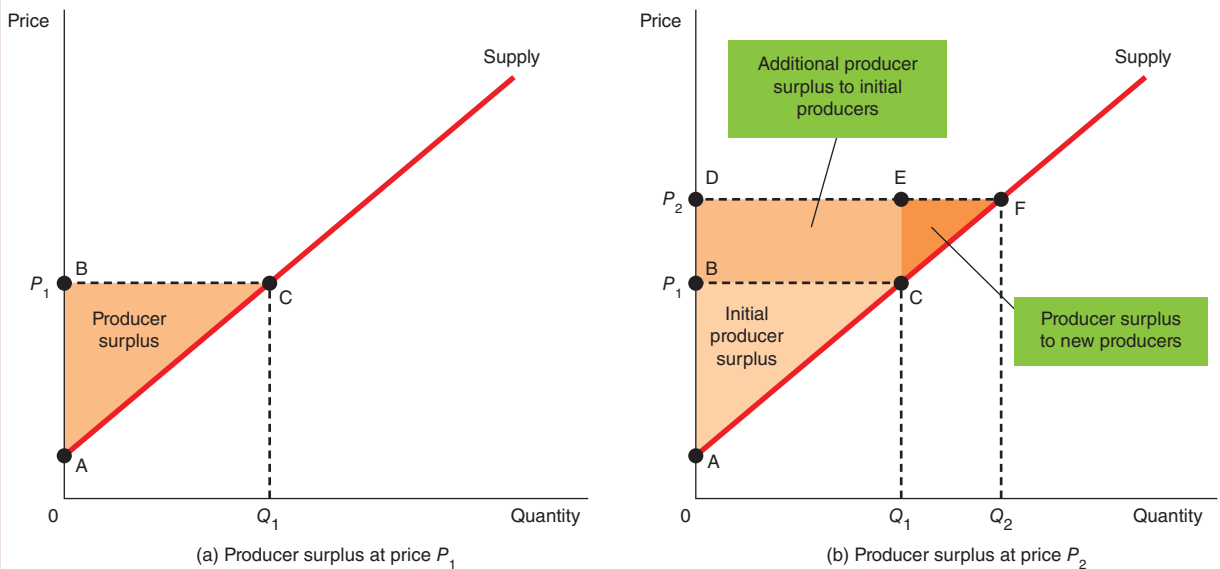
You will not be surprised to hear that sellers always want to receive a higher price for the goods they sell. But how much does sellers' well-being rise in response to a higher price? The concept of producer surplus offers a precise answer to this question.

Figure 7.6 shows a typical upwards sloping supply curve. Even though this supply curve differs in shape from the step-like supply curves in the previous figure, we measure producer surplus in the same way: producer surplus is the area below the price and above the supply curve. In panel (a), the price is P_1 and producer surplus is the area of triangle ABC.

FIGURE 7.6

How the Price Affects Producer Surplus

In panel (a) the price is P_1 , the quantity demanded is Q_1 and producer surplus equals the area of the triangle ABC. When the price rises from P_1 to P_2 , as in panel (b), the quantity supplied rises from Q_1 to Q_2 and the producer surplus rises to the area of the triangle ADF. The increase in producer surplus (area BCFD) occurs in part because existing producers now receive more (area BCED) and in part because new producers enter the market at the higher price (area CEF).



SELF TEST Is producer surplus the same as profit? Explain.

If the demand for a good fell, what would happen to price and the producer surplus in the market as a result?

Use an appropriate supply and demand diagram to help support your answer.

Panel (b) shows what happens when the price rises from P_1 to P_2 . Producer surplus now equals area ADF. This increase in producer surplus has two parts. First, those sellers who were already selling Q_1 of the good at the lower price P_1 are better off because they now get more producer surplus for what they sell. The increase in producer surplus for existing sellers equals the area of the rectangle BCED. Second, some new sellers enter the market because they are now willing to produce the good at the higher price, resulting in an increase in the quantity supplied from Q_1 to Q_2 . The producer surplus of these newcomers is the area of the triangle CEF.

MARKET EFFICIENCY

In economies throughout the world, there are trades made every day – millions of them. As part of our analysis of free markets we make many assumptions. For example, we assume that if an individual gives up €25 to buy a pair of jeans, the value of the economic benefit that individual gains is at the very least equal to the price paid but may include some consumer surplus if they were prepared to pay more than €25 to buy those jeans. Equally, we are assuming that the seller must value the sale to the consumer at the very least at €25 or they would not have sold the jeans and indeed may be getting some producer surplus from the sale.

Can we conclude, therefore, that the allocative outcome in a free market is efficient? To do this we have to define efficiency in this context.

Consumer and producer surplus provide a way in which we can measure the benefits to consumers and producers of trading. Remember that in Chapter 5, the consumer optimum was the point where the marginal rate of substitution of an extra euro spent equalled the marginal utility of that extra euro. In Chapter 6, the optimum for a profit-maximizing producer was the point where the marginal cost of an extra unit produced equalled the marginal revenue of an extra unit. In both cases we have noted that there will be incentives for firms and consumers to change their behaviour if they are at any point other than their respective equilibriums.

We have also noted that Adam Smith's theory of the invisible hand suggests that millions of independent decision makers, both consumers and producers, all go about their business but market forces lead to a degree of coherence between these decisions. In theory, free market economies will not tend to have instances where there are vast shortages and surpluses for long periods of time because there will be incentives for producers and consumers to change their behaviour which moves the market to equilibrium.

This analysis is the basis for what is called general equilibrium. **General equilibrium** is the notion that the decisions and choices of economic agents are coordinated across markets. General equilibrium deals with the idea that the market mechanism leads to outcomes that are efficient. Consumers are maximizing utility and producers are maximizing profits and producing at minimum average cost. In more advanced courses you will study the theory of the relationship between consumer theory and producer theory.

general equilibrium the notion that the decisions and choices of economic agents are coordinated across markets

For the moment we are going to look at the notion of economic efficiency in relation to the analysis of consumer and producer surplus presented so far.

Economic Efficiency and Waste

If we were to look at efficiency as a general concept we would be likely to introduce the word 'waste'; if something is inefficient it is wasteful. To the consumer, spending money on a good which does not provide

value can be considered a waste. Equally, if a producer spends money on producing a good which consumers do not want to buy, or on a combination of factors that could be re-organized differently to reduce costs, then that also represents waste.

We can ask the question, therefore, are free markets wasteful? If waste existed then there would be a way to reallocate resources to reduce that waste – consumers would adjust their buying habits and producers their production methods. Consumer surplus is the benefit that buyers receive from participating in a market, and producer surplus is the benefit that sellers receive. At any point on the demand curve, therefore, the price represents the value placed on the good by consumers on the last unit consumed, whereas any point on the supply curve represents the additional cost to a producer in the market of producing one more unit. At any point we can also measure the consumer and producer surplus that exists. If at this particular price the quantity demanded is higher than the quantity supplied it tells us that the value placed on the additional unit by consumers is higher than the additional cost to the producer. In market equilibrium, therefore, the value of the additional unit to buyers is the same as the additional cost to producers. We can look at the consumer and producer surplus at equilibrium and add these together to get a measure of the **total surplus**. If the consumer surplus is a measure of the consumer's well-being and producer surplus is a measure of the seller's well-being, then total surplus can be used as a measure of society's economic well-being. We can summarize this as:

$$\text{Total surplus} = \text{Value to buyers} - \text{Cost to sellers}$$

total surplus the total value to buyers of the goods, as measured by their willingness to pay, minus the cost to sellers of providing those goods

If an allocation of resources maximizes total surplus, we say that the allocation exhibits **efficiency**. If an allocation is not efficient, then some of the gains from trade among buyers and sellers are not being realized. For example, an allocation is inefficient if a good is not being produced by the sellers with lowest cost. In this case, moving production from a high-cost producer to a low-cost producer will lower the total cost to sellers and raise total surplus. Similarly, an allocation is inefficient if a good is not being consumed by the buyers who value it most highly. In this case, moving consumption of the good from a buyer with a low valuation to a buyer with a high valuation will raise total surplus. In Chapter 1, we defined *efficiency* as 'the property of society getting the most it can from its scarce resources'. Now that we have the concept of total surplus, we can be more precise about what we mean by 'getting the most it can'. In this context, society will be getting the most it can from its scarce resources if it allocates them so as to maximize total surplus.

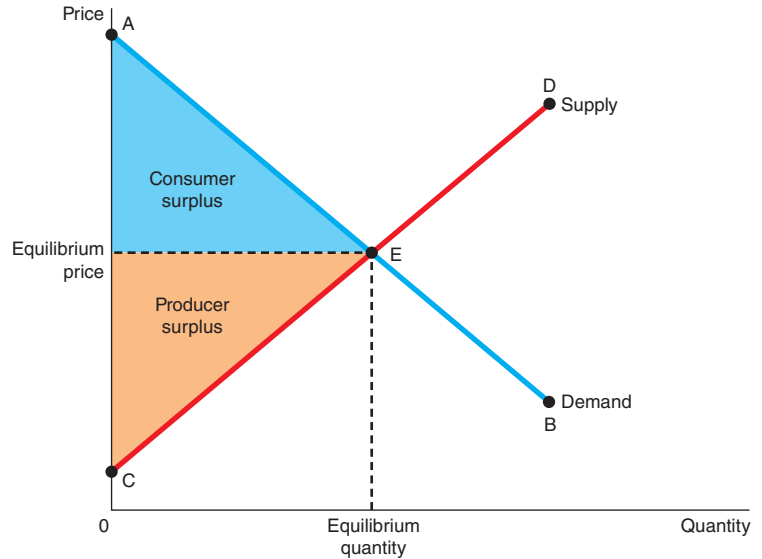
efficiency the property of a resource allocation of maximizing the total surplus received by all members of society

Figure 7.7 shows consumer and producer surplus when a market reaches the equilibrium of supply and demand. Recall that consumer surplus equals the area above the price and under the demand curve and producer surplus equals the area below the price and above the supply curve. Thus, the total area between the supply and demand curves up to the point of equilibrium represents the total surplus in this market.

Pareto Efficiency At this point we will introduce the concept of Pareto efficiency. The idea was developed by an Italian economist, Wilfredo Pareto (1848–1923). Pareto efficiency occurs if it is not possible to reallocate resources in such a way as to make one person better off without making anyone else worse off. Markets are all about trading and, as we have seen, the demand curve tells us something about the benefit consumers receive from allocating their income in a particular way, and the supply curve the benefit to suppliers of offering goods for sale. When trades take place, the consumer gains some benefit and so does the producer and this is referred to as a **Pareto improvement**. A Pareto improvement occurs when an action makes at least one economic agent better off without harming another economic agent.

FIGURE 7.7**Consumer and Producer Surplus in the Market Equilibrium**

Total surplus – the sum of consumer and producer surplus – is the area between the supply and demand curves up to the equilibrium quantity.



Pareto improvement when an action makes at least one economic agent better off without harming another economic agent

Consumers and producers, therefore, will continue to readjust their decision making with the resulting reallocation of resources until any Pareto improvements are exhausted. We can view economic efficiency, therefore, in terms of the point where all possible Pareto improvements have been exhausted.

Evaluating the Market Equilibrium

We have noted that total surplus is maximized at the point where the market is in equilibrium – an allocation of resources where consumers are maximizing utility and producers maximizing profits and producing at minimum average cost. Is it possible to reallocate resources in any other way to increase the well-being of consumers and producers, in other words, are there Pareto improvements that would result from any such resource allocations?

The price determines which buyers and sellers participate in the market. Those buyers who value the good more than the price (represented by the segment AE on the demand curve in Figure 7.7) choose to buy the good; those buyers who value it less than the price (represented by the segment EB) do not. Similarly, those sellers whose costs are less than the price (represented by the segment CE on the supply curve) choose to produce and sell the good; those sellers whose costs are greater than the price (represented by the segment ED) do not.

These observations lead to two insights about market outcomes:

1. Free markets allocate the supply of goods to the buyers who value them most highly, as measured by their willingness to pay.
2. Free markets allocate the demand for goods to the sellers who can produce them at least cost.

Thus, given the quantity produced and sold in market equilibrium, economic well-being cannot be increased by consumers or producers changing their respective allocations.

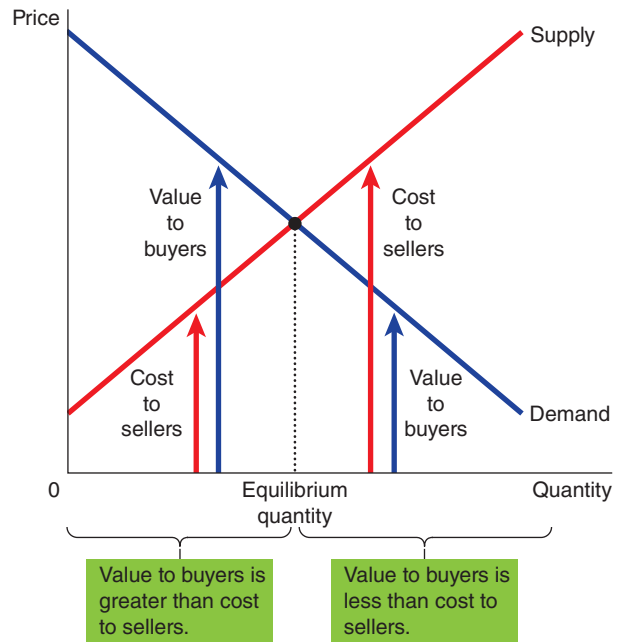
We can also identify a third insight about market outcomes:

3. Free markets produce the quantity of goods that maximizes the sum of consumer and producer surplus.

To see why this is true, consider Figure 7.8. Recall that the demand curve reflects the value to buyers and that the supply curve reflects the cost to sellers. At quantities below the equilibrium level, the value to buyers exceeds the cost to sellers. In this region, increasing the quantity raises total surplus, and it continues to do so until the quantity reaches the equilibrium level. Beyond the equilibrium quantity, however, the value to buyers is less than the cost to sellers. Producing more than the equilibrium quantity would, therefore, lower total surplus.

FIGURE 7.8**The Efficiency of the Equilibrium Quantity**

At quantities less than the equilibrium quantity, the value to buyers exceeds the cost to sellers. At quantities greater than the equilibrium quantity, the cost to sellers exceeds the value to buyers. Therefore, the market equilibrium maximizes the sum of producer and consumer surplus.



These three insights about market outcomes tell us that the equilibrium outcome is an efficient allocation of resources. This conclusion explains why some economists advocate free markets as a preferred way to organize economic activity.

SELF TEST Look back at the section on consumer optimum from Chapter 5 and on producer equilibrium from Chapter 6. How might you explain the relationship between these equilibrium points and market equilibrium in terms of economic efficiency?

Efficiency and Equity

Much of this chapter has focused on efficiency. It is not surprising that there is a great deal of focus on efficiency in economics because it is something that can be measured and is quantifiable. Efficiency is a positive concept in that it can be stated *what is* an efficient allocation. This does not, however, tell us anything about whether the efficient allocation is desirable or not.

We also have to consider whether an allocation is fair and this is a normative concept. One way of looking at fairness in economic allocations is to consider **equity** – the property of distributing economic prosperity fairly among the members of society. In essence, the gains from trade in a market are like a cake to be distributed among the market participants. The question of efficiency is whether the cake is

as big as possible. The question of equity is whether the cake is divided fairly. Evaluating the equity of a market outcome is more difficult than evaluating the efficiency. Whereas efficiency is an objective goal that can be judged on strictly positive grounds, equity involves normative judgements that go beyond economics and enter into the realm of political philosophy.

equity the property of distributing economic prosperity fairly among the members of society

One of the problems with the analysis we have presented is an assumption that economic agents are all similar – that consumers and producers are a heterogeneous group. Clearly this is not the case. One of the most important things economists have to consider is the different way that people with different income endowments behave. The marginal utility gained from spending an extra unit of income for a very poor person is likely to be very different to that of a rich person, for example. Looking at well-being simply from the perspective of adding up the consumer and producer surpluses does hide more complex issues.

Some economists point to the collective utility of society which is reflected by consumer and producer surplus in terms of a **social welfare function**. Social welfare functions attempt to take into account the fact that the marginal utilities of individual households are not all the same and indeed that their preferences are also different. This is based on the assumption that welfare is an ordinal function, i.e. that consumers can rank preferences. However, it is also assumed that households operate with imperfect knowledge. Decisions in the market may be made by those who have some power which can distort market outcomes, for example the spending power of the rich or those with political influence can mean that market outcomes are disproportionately skewed. The outcome may be efficient but it is not necessarily fair.

social welfare function the collective utility of society which is reflected by consumer and producer surplus

We must take into account, therefore, that different stakeholders will have different perspectives on what is ‘good’ for society as a whole. One example of these different perspectives, which are referred to as social states, is in relation to income – would you rather have a society that focused on raising total income regardless of how it was distributed, i.e. that a small number of people owned a considerable portion of this income, or a social state where income was more evenly distributed among its citizens?

In the next chapters we begin to look at how these issues take on more relevance when governments get involved in free markets by affecting market outcomes to improve them. Clearly, if a government says it is adopting policies to try and improve market outcomes we are in the realm of normative economics and looking at *what should* be a market outcome rather than *what is*.

CONCLUSION: MARKET EFFICIENCY AND MARKET FAILURE

This chapter introduced the basic tools of welfare economics – consumer and producer surplus – and used them to evaluate the efficiency of free markets. We showed that the forces of supply and demand allocate resources efficiently. That is, even though each buyer and seller in a market is concerned only about his or her own welfare, they are together led by an invisible hand to an equilibrium that maximizes the total benefits to buyers and sellers.

A word of warning is in order. To conclude that markets are efficient, we made several assumptions about how markets work. When these assumptions do not hold, our conclusion that the market equilibrium is efficient may no longer be true. In addition, a market outcome may be identified as efficient but it does not follow that this particular outcome is fair.

IN THE NEWS

Well-Being

Welfare economics has been a sub-topic in the discipline since the 19th century but exactly how we measure well-being is something that has been repeatedly visited and refined over the years. In the UK, the Office for National Statistics (ONS) established the Measuring National Well-being programme in 2010 and in 2012 published the first annual report on Life in the UK (see http://www.ons.gov.uk/ons/dcp171766_287415.pdf).

Measuring National Well-being: Life in the UK, 2012

2012 was not the best year for many people in the UK. The economy slipped into its second recession in four years as problems with the Eurozone and government policies designed to cut the budget deficit, coupled with an inflation rate that was higher than the government's 2.0 per cent target, meaning many households saw their real incomes fall. In such circumstances, surely well-being would fall?

One of the first problems is how you measure well-being. In this chapter we have looked at using consumer and producer surplus as a measure of welfare but also noted that such a measure relies on a set of assumptions that many question

and also needs to take into account notions of equity. The ONS has decided to look at a range of factors which include individual perceptions of well-being, the state of the environment, sustainability, economic performance, quality of life and equality, to gauge national well-being.

The results of the survey conducted by the ONS reveal some interesting observations. In 2010/11, the ONS noted that 21 per cent of people lived in poverty, defined by if household income is less than 60 per cent of the median (middle) net household income before and after housing costs. Despite this only 5.8 per cent of people surveyed said that they were 'completely dissatisfied' with their income. When looking at wealth, which includes ownership

of assets such as property, pensions and share ownership, 12.3 per cent of adults over 16 said they found it quite difficult to manage financially.

The ONS state that an important component of national well-being is how individuals perceive their well-being – how they think and feel about their own lives. Of the people surveyed, 75.9 per cent rated their life with a medium to high rating of satisfaction with life overall, 80 per cent said they felt the things they do are worthwhile and 71.1 per cent rated the question 'how happy were you yesterday?' being medium to high. This presents a fairly positive picture of UK citizens' well-being but the ONS also points out that there were variations in the answers received dependent on age, health, relationship, employment status and location.

Whilst employment provides income to spend on necessities which can affect perceptions of well-being, spending on leisure activities is also an important aspect of well-being. It is not simply having the money to access leisure but also the time. Almost a quarter of people over 16 surveyed (24.8 per cent) said that they were somewhat, mostly or completely dissatisfied with the amount of leisure time they have, whilst 62.3 per cent said that they were somewhat, mostly or completely satisfied. The remainder (12.9 per cent) were neither satisfied nor dissatisfied.

People in any country are affected to a considerable degree

Poverty and wealth often cohabit side-by-side but do riches make people happier?



by the actions of government and an interesting result of the survey showed that trust in the government has fallen from 41 per cent in 2007 to 17 per cent in 2009 and was 23 per cent in 2012. Reasons for this decline in the 'tending to trust' category might be due to a variety of reasons which include the economic difficulties facing the UK, to having a coalition government and various high profile scandals.

Questions

- 1 Are you surprised that despite the problems facing UK citizens in 2012 the individual perceptions of well-being are relatively high? Explain.
- 2 How do the measures of well-being used by the ONS differ from that which we have described in this chapter? Which do you think is the better measure and why?
- 3 Would you have expected more people living in poverty to be 'completely dissatisfied' with their incomes? What factors might explain this relatively small proportion answering the question in this way?
- 4 Consider the methodology used by the ONS in the survey in asking people to respond to questions through a scale which asks them to answer from 'completely satisfied' to 'completely dissatisfied'. What sort of scale is this and why might it present problems in interpreting people's responses?
- 5 Do you think economists should focus on measuring efficiency, which is a positive concept, and pay less attention to equity, which is a normative concept? How might research like the ONS 'Life in the UK, 2012' survey help in the analysis of well-being?

SUMMARY

- Consumer surplus equals buyers' willingness to pay for a good minus the amount they actually pay for it, and it measures the benefit buyers get from participating in a market. Consumer surplus can be computed by finding the area below the demand curve and above the price.
- Producer surplus equals the amount sellers receive for their goods minus their costs of production, and it measures the benefit sellers get from participating in a market. Producer surplus can be computed by finding the area below the price and above the supply curve.
- An allocation of resources that maximizes the sum of consumer and producer surplus is said to be efficient. Policymakers are often concerned with the efficiency, as well as the equity, of economic outcomes.
- The equilibrium of supply and demand maximizes the sum of consumer and producer surplus. That is, the invisible hand of the marketplace leads buyers and sellers to allocate resources efficiently.

QUESTIONS FOR REVIEW

- 1 What is meant by the term 'allocative efficiency'?
- 2 What is welfare economics?
- 3 Explain how buyers' willingness to pay, consumer surplus and the demand curve are related.
- 4 Explain how sellers' costs, producer surplus and the supply curve are related.
- 5 Prepare a supply-and-demand diagram, showing producer and consumer surplus in the market equilibrium.
- 6 What is efficiency and how might we measure it?
- 7 Using the supply-and-demand diagram you drew for Question 5, assume that the demand shifts to the right as a result of an increase in incomes. On the diagram, show how consumer surplus and producer surplus changes as a result of the shift in demand. Is total surplus increased, decreased or does it stay the same? What would the outcome depend on?
- 8 What is meant by a Pareto efficient outcome?
- 9 Why are issues relating to efficiency classed as positive? What are the normative issues we might want to be concerned with?
- 10 Explain how charging a higher price for tickets for the Wimbledon tennis championships would not only get rid of queues but also would be a more efficient market allocation. Would this also be an equitable market allocation? Explain.

PROBLEMS AND APPLICATIONS

- How would you define the concept of *welfare*? If well-being is about happiness and satisfaction with life, how would you, as an economist, go about trying to define and quantify happiness and satisfaction with life? Is it possible to do so?
- An early freeze in Normandy ruins half of the apple harvest. What happens to consumer surplus in the market for apples? What happens to consumer surplus in the market for cider? Illustrate your answers with diagrams.
- Suppose the demand for French bread rises. What happens to producer surplus in the market for French bread? What happens to producer surplus in the market for flour? Illustrate your answer with diagrams.
- It is a hot day, and Günter is thirsty. Here is the value he places on a bottle of water:

| | |
|------------------------|----|
| Value of first bottle | €7 |
| Value of second bottle | €5 |
| Value of third bottle | €3 |
| Value of fourth bottle | €1 |

- From this information, derive Günter's demand schedule. Graph his demand curve for bottled water.
 - If the price of a bottle of water is €4, how many bottles does Günter buy? How much consumer surplus does Günter get from his purchases? Show Günter's consumer surplus in your graph.
 - If the price falls to €2, how does quantity demanded change? How does Günter's consumer surplus change? Show these changes in your graph.
- Mario owns a water pump. Because pumping large amounts of water is harder than pumping small amounts, the cost of producing a bottle of water rises as he pumps more. Here is the cost he incurs to produce each bottle of water:

| | |
|-----------------------|----|
| Cost of first bottle | €1 |
| Cost of second bottle | €3 |
| Cost of third bottle | €5 |
| Cost of fourth bottle | €7 |

- From this information, derive Mario's supply schedule. Graph his supply curve for bottled water.
 - If the price of a bottle of water is €4, how many bottles does Mario produce and sell? How much producer surplus does Mario get from these sales? Show Mario's producer surplus in your graph.
 - If the price rises to €6, how does quantity supplied change? How does Mario's producer surplus change? Show these changes in your graph.
- Consider a market in which Günter from Problem 4 is the buyer and Mario from Problem 5 is the seller.
 - Use Mario's supply schedule and Günter's demand schedule to find the quantity supplied and quantity

- demand at prices of €2, €4 and €6. Which of these prices brings supply and demand into equilibrium?
 - What are consumer surplus, producer surplus and total surplus in this equilibrium?
 - If Mario produced and Günter consumed one fewer bottle of water, what would happen to total surplus?
 - If Mario produced and Günter consumed one additional bottle of water, what would happen to total surplus?
- Why might we want to think about market price as the outcome of a bargaining model?
 - The cost of producing smart phones has fallen over the past few years. Let's consider some implications of this fact.
 - Use a supply-and-demand diagram to show the effect of falling production costs on the price and quantity of smart phones sold.
 - In your diagram, show what happens to consumer surplus and producer surplus.
 - Suppose the supply of smart phones is very elastic. Who benefits most from falling production costs – consumers or producers of smartphones?

- Four consumers are willing to pay the following amounts for haircuts:

| | | | |
|--------------|--------------|---------------|----------------|
| Hans: | Juan: | Peter: | Marcel: |
| €7 | €2 | €8 | €5 |

There are four haircutting businesses with the following costs:

| | | | |
|----------------|----------------|----------------|----------------|
| Firm A: | Firm B: | Firm C: | Firm D: |
| €3 | €6 | €4 | €2 |

Each firm has the capacity to produce only one haircut. For efficiency, how many haircuts should be given? Which businesses should cut hair, and which consumers should have their hair cut? How large is the maximum possible total surplus?

- Suppose a technological advance reduces the cost of making tablet devices.
 - Use a supply and demand diagram to show what happens to price, quantity, consumer surplus and producer surplus in the market for tablet devices.
 - Tablet devices and laptops are substitutes. Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus and producer surplus in the market for laptops. Should laptop producers be happy or sad about the technological advance in tablet devices?
 - Tablet devices and apps are complements. Use a supply-and-demand diagram to show what happens to price, quantity, consumer surplus and producer surplus in the market for apps. Should app producers be happy or sad about the technological advance in tablet devices?

8

SUPPLY, DEMAND AND GOVERNMENT POLICIES

In the last four chapters we looked at how markets work under the assumptions of perfect competition. When these assumptions do not hold, our conclusion that the market equilibrium is efficient may no longer be true. In the real world, competition is sometimes far from perfect. In order to bring about a more equitable outcome, governments develop policies which are designed to improve market outcomes. In this chapter we will analyse various types of government policy using only the tools of supply and demand.

Policies often have effects that their architects did not intend or anticipate. One of the *Ten Principles of Economics* is that people respond to incentives. When new policies are introduced they may have good intentions to solve particular problems or help groups of people but new policies also generate new incentives and as a result the consequences may go far beyond those envisaged by the policymakers.

We begin by considering policies that directly control prices which are usually enacted when policymakers believe that the market price of a good or service is unfair to buyers or sellers. Yet, as we will see, these policies can generate inequities of their own. We will then consider the impact of taxes and subsidies. Policymakers use taxes and subsidies to influence market outcomes and, in the case of taxes, to raise revenue for public purposes.

CONTROLS ON PRICES

We will look at two policies to control prices – price ceilings and price floors. These policies may be introduced by a government or a regulatory body but in some cases might also be set by a business, for example prices are often set by sports and entertainment bodies which have similar results to legal price controls.

A **price ceiling** or a price cap is a legal maximum on the price at which a good can be sold. A **price floor** is the exact opposite – a minimum price that producers can charge for a good; producers are not allowed to charge a price any lower than the legal minimum. Let us consider the effects of these policies in turn.

price ceiling a legal maximum on the price at which a good can be sold
price floor a legal minimum on the price at which a good can be sold

To see how price controls affect market outcomes, we will look at an example which has interested economists for many years – rent control. Many people around Europe do not own their own homes and instead choose to rent one. If rental space (we will define this in terms of square metres) is sold in a competitive market free of government regulation, the price of housing adjusts to balance supply and demand: at the equilibrium price, the quantity of housing that buyers want to buy exactly equals the quantity that sellers want to sell. To be concrete, suppose the equilibrium price is €30m².

Not everyone may be happy with the outcome of this free market process. People looking for homes might complain that the €30m² is too high for those seeking a home to afford. Landlords, on the other hand, might complain that €30m² is the result of 'cut-throat competition', is too low and is depressing the incomes of landlords. House seekers and landlords lobby the government to pass laws that alter the market outcome by directly controlling the price of rental accommodation.

If those seeking rental accommodation are successful in their lobbying, the government imposes a legal maximum on the price at which rental accommodation can be sold. Let us consider the effects of this policy.

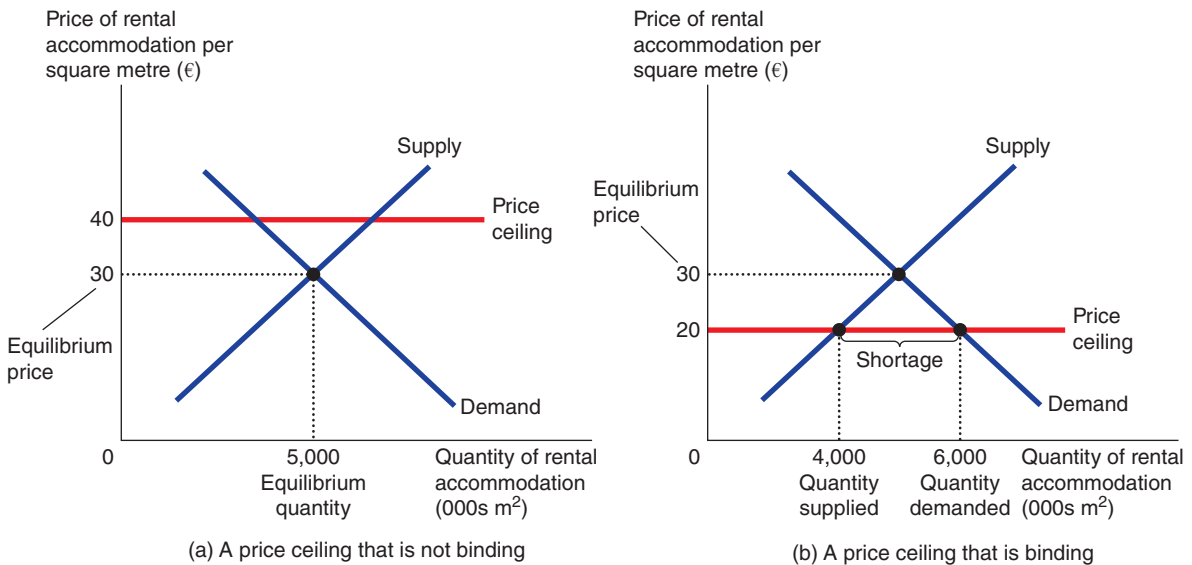
How Price Ceilings Affect Market Outcomes

When a price ceiling is imposed, two outcomes are possible. In panel (a) of Figure 8.1, the government imposes a price ceiling of €40m². In this case, because the price that balances supply and demand (€30m²) is below the ceiling, the price ceiling is *not binding*. Market forces naturally move the economy to the equilibrium, and the price ceiling has no effect on the price or the quantity sold.

FIGURE 8.1

A Market With a Price Ceiling

In panel (a), the government imposes a price ceiling of €40m². Because the price ceiling is above the equilibrium price of €30m², the price ceiling has no effect, and the market can reach the equilibrium of supply and demand. In this equilibrium, quantity supplied and quantity demanded both equal 5 million square feet. In panel (b), the government imposes a price ceiling of €20m². Because the price ceiling is below the equilibrium price of €30m², the market price equals €20m². At this price, 6 million square feet is demanded and only 4 million square feet is supplied, so there is a shortage of 2 million square feet.



Panel (b) of Figure 8.1 shows the other, more interesting, possibility. In this case, the government imposes a price ceiling of €20m². Because the equilibrium price of €30m² is above the price ceiling, the ceiling is a *binding constraint* on the market. The forces of supply and demand tend to move the price towards the equilibrium price, but when the market price hits the ceiling, it can rise no further. Thus, the market price equals the price ceiling. Given this binding limit on price, incentives change. Some landlords will not find it profitable to rent out property at this price and remove the property from the market. Buyers' incentives are different. The lower price of rental accommodation means the sacrifice they have to make is less in terms of alternatives foregone and so are prepared to buy more at the binding price. At this price, the quantity of rental accommodation demanded (6 million square metres in Figure 8.1) exceeds the quantity supplied (4 million square metres). There is a shortage of rental accommodation, so some people who want to rent accommodation at the going price are unable to.

When a shortage of rental accommodation develops because of this price ceiling, some mechanism for rationing accommodation will naturally develop. The mechanism could simply be long queues: buyers who

are willing to arrive early when rental accommodation becomes available and queue, get a home, while those unwilling to wait do not. Alternatively, sellers could ration accommodation according to their own personal biases, selling it only to friends, relatives, or members of their own racial or ethnic group. Notice that even though the price ceiling was motivated by a desire to help those seeking to rent homes, not all buyers benefit from the policy. Some buyers do get to pay a lower price, although they may have to queue a long time to do so, but other buyers cannot get any accommodation at all.

In setting policy, the authorities have to consider who gets the benefits and the possible value of these benefits in relation to the cost. The benefit in this case are those buyers who are able to rent accommodation at €20m², whereas the cost are those who want to rent accommodation but are unable to do so. There might also be costs in terms of the effect on sellers that has to be taken into consideration. Some sellers may not feel it is worth their while continuing in the market and leave, thus depressing market supply. There will also be the effect on workers, such as construction workers involved with repair and maintenance of rental property and letting agents, who may be made redundant as their services are no longer required. To society as a whole there is an opportunity cost of these redundant factor inputs.

This example in the market for rental accommodation shows a general result: when the government imposes a binding price ceiling on a competitive market, a shortage of the good arises, and sellers must ration the scarce goods among the large number of potential buyers. The rationing mechanisms that develop under price ceilings are rarely desirable. Long queues are inefficient because they waste buyers' time. (The queue may not be a physical one; people may be on a waiting list hoping to hear that the good has become available.) Discrimination according to seller bias is both inefficient (because the good does not necessarily go to the buyer who values it most highly) and potentially unfair. By contrast, the rationing mechanism in a free, competitive market is both efficient and impersonal. When the market for rental accommodation reaches its equilibrium, anyone who wants to pay the market price can get accommodation. Free markets ration goods with prices although as we have seen, whilst this outcome might be efficient it might not be equitable.

CASE STUDY Rent Control in the Short Run and Long Run

Economists often criticize rent control, arguing that it is a highly inefficient way to help the poor raise their standard of living. One economist called rent control 'the best way to destroy a city, other than bombing'.

The adverse effects of rent control are less apparent to the general population because these effects occur over many years. In the short run, landlords have a fixed number of housing units (houses and flats) to rent, and they cannot adjust this number quickly as market conditions change. Moreover, the number of people searching for housing in a city may not be highly responsive to rents in the short run because people take time to adjust their housing arrangements. Therefore, the short-run supply and demand for housing are relatively inelastic.

Panel (a) of Figure 8.2 shows the short-run effects of rent control on the housing market. As with any binding price ceiling, rent control causes a shortage. Yet because supply and demand are inelastic in the short run, the initial shortage caused by rent control is small. The primary effect in the short run is to reduce rents.

The long-run story is very different because the buyers and sellers of rental housing respond more to market conditions as time passes – there is an incentive for both buyers and sellers to change their behaviour in response to the changed market conditions. On the supply side, landlords respond to low rents by not building new housing and by failing to maintain existing housing. Some are likely to feel that they are not able to generate any return on investing in improving or maintaining properties because they cannot charge higher prices to compensate for the costs they have incurred. On the demand side, low rents encourage people to find their own housing (rather than living with their parents or sharing with friends) and induce more people to move into a city. Therefore, both supply and demand are more elastic in the long run.

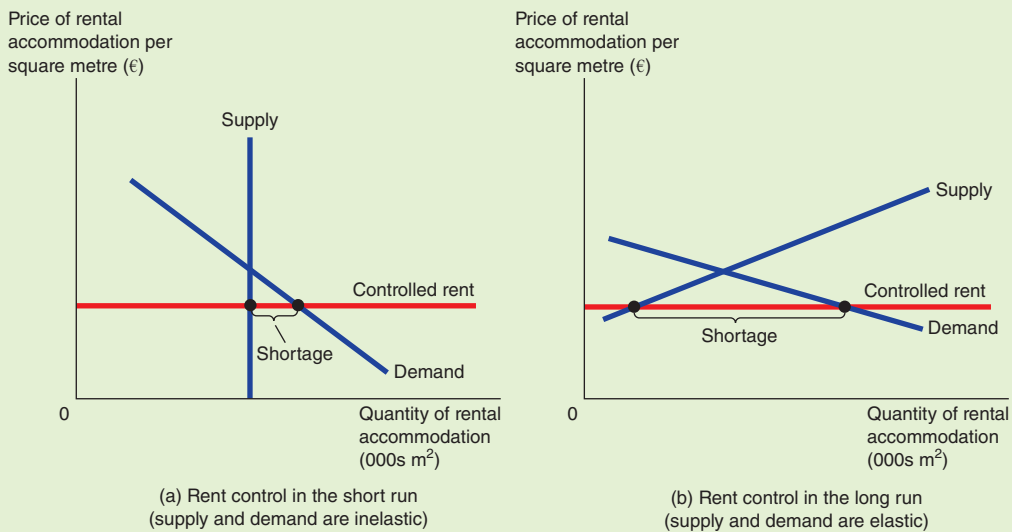
Panel (b) of Figure 8.2 illustrates the housing market in the long run. When rent control depresses rents below the equilibrium level, the quantity of housing units supplied falls substantially, and the quantity of housing units demanded rises substantially. The result is a large shortage of housing.

(Continued)

In cities with rent control, landlords use various mechanisms to ration housing. Some landlords keep long waiting lists. Others give a preference to tenants without children. Still others discriminate on the basis of race. Sometimes, apartments are allocated to those willing to offer under-the-table payments – the so-called ‘black market’. In essence, these bribes bring the total price of an apartment (including the bribe) closer to the equilibrium price.

FIGURE 8.2**Rent Control in the Short Run and in the Long Run**

Panel (a) shows the short-run effects of rent control: because the supply and demand for housing units are relatively inelastic, the price ceiling imposed by a rent-control law causes only a small shortage of housing. Panel (b) shows the long-run effects of rent control: because the supply and demand for housing units are more elastic, rent control causes a large shortage.



Rent control provides an excellent example of Principle 4 of the *Ten Principles of Economics*: people respond to incentives. In free markets, landlords try to keep their buildings clean and safe because desirable apartments command higher prices. By contrast, when rent control creates shortages and waiting lists, landlords lose their incentive to respond to tenants' concerns. Why should a landlord spend his money to maintain and improve his property when people are waiting to get in as it is? In the end, tenants get lower rents, but they also get lower-quality housing.

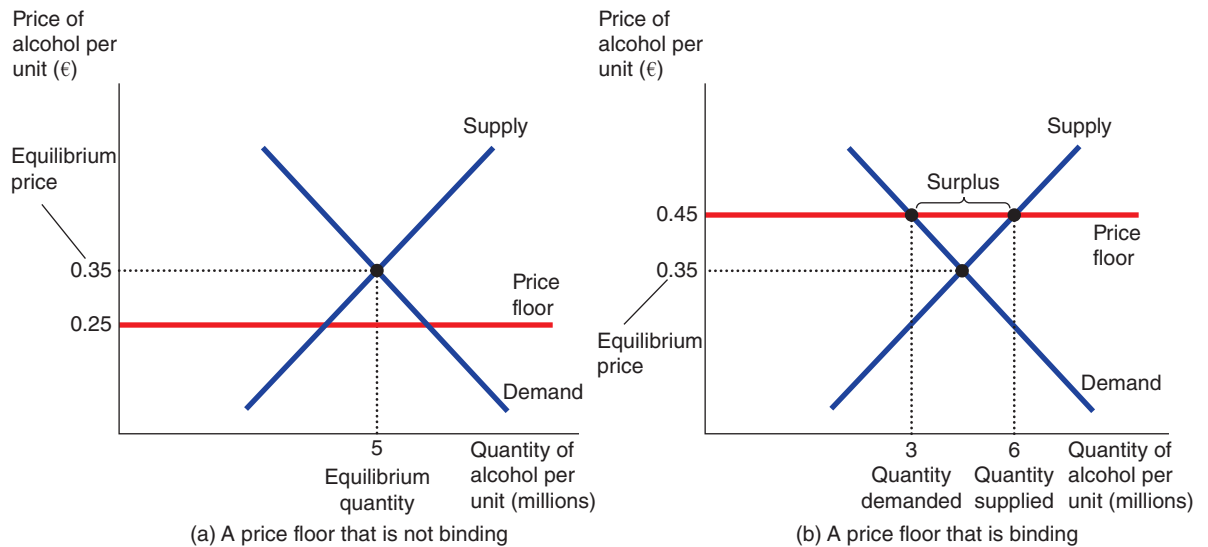
Policymakers often react to the effects of rent control by imposing additional regulations. For example, there are laws that make racial discrimination in housing illegal and require landlords to provide minimally adequate living conditions. These laws, however, are difficult and costly to enforce. By contrast, when rent control is eliminated and a market for housing is regulated by the forces of competition, such laws are less necessary. In a free market, the price of housing adjusts to eliminate the shortages that give rise to undesirable landlord behaviour.

How Price Floors Affect Market Outcomes

To examine the effects of another kind of government price control, let's look at the market for alcohol. In England and Scotland, the governments have been planning policies to set minimum prices for alcohol to try and curb the damaging effects of excess alcohol consumption on both their citizen's health and on social behaviour. Price floors, like price ceilings, are an attempt by the government to maintain prices at other than equilibrium levels.

FIGURE 8.3**A Market with a Price Floor**

In panel (a), the government imposes a price floor of €0.25 per unit. Because this is below the equilibrium price of €0.35 per unit, the price floor has no effect. The market price adjusts to balance supply and demand. At the equilibrium, quantity supplied and quantity demanded both equal 5 million units. In panel (b), the government imposes a price floor of €0.45 per unit, which is above the equilibrium price of €0.35 per unit. Therefore, the market price equals €0.45 per unit. Because 6 million units are supplied at this price and only 3 million are demanded, there is a surplus of 3 million units.



When the government imposes a price floor on the market for alcohol, two outcomes are possible. If the government imposes a price floor of €0.25 per unit when the equilibrium price is €0.35, we obtain the outcome in panel (a) of Figure 8.3. In this case, because the equilibrium price is above the floor, the price floor is not binding. Market forces naturally move the economy to the equilibrium, and the price floor has no effect.

Panel (b) of Figure 8.3 shows what happens when the government imposes a price floor of €0.45 per unit. In this case, because the equilibrium price of €0.35 per unit is below the floor, the price floor is a binding constraint on the market. The forces of supply and demand tend to move the price towards the equilibrium price, but when the market price hits the floor, it can fall no further. The market price equals the price floor. At this floor, the quantity of alcohol supplied, 6 million units, exceeds the quantity demanded (3 million units). Some people who want to sell alcohol at the going price are unable to. Thus, a binding price floor causes a surplus. Notice, however, that the price floor has reduced the quantity demanded by 2 million units, from the original equilibrium quantity demanded of 5 million units to 3 million units, thus achieving the goal of the policy.

Just as price ceilings and shortages can lead to undesirable rationing mechanisms, so can price floors and surpluses. In the case of a price floor, some sellers are unable to sell all they want at the market price. The sellers who appeal to the personal biases of the buyers, perhaps due to income levels, for example, are better able to sell their goods than those who do not. The poor might be disproportionately affected by the higher price for alcohol whilst those people who would see themselves as responsible drinkers have to pay a higher price because of the proportion of drinkers who do abuse alcohol in some way and impose costs on society. By contrast, in a free market, the price serves as the rationing mechanism, and sellers can sell all they want at the equilibrium price.

Summary

Price controls are used when governments or other agencies believe that the market is not allocating resources equitably (even if it is allocating resources efficiently). Just as there is an argument that an efficient allocation of resources is not an equitable one, there are costs and benefits of imposing price controls and there is considerable debate as to whether price controls are desirable or not.

There are other options available to governments for achieving what might be seen as being equitable outcomes and we shall look at these in more detail in the next section when we look at taxes and subsidies.

SELF TEST How might price ceilings like rent controls lead to those where the policy is designed to help, being negatively affected?
What might be the costs to individuals of price floors?

TAXES

Most governments, whether national or local, impose taxes to raise revenue and influence behaviour in some way. Because taxes are such an important policy instrument, and because they affect our lives in many ways, the study of taxes is a topic to which we return several times throughout this book. In this section we begin our study of how taxes affect the economy.

There are many different taxes in most countries but we can generally divide taxes into two categories: taxes on income and taxes on spending. Taxes on income are called **direct taxes** because the individual is ultimately responsible for paying the correct amount of tax. For many people, their monthly salary payment is adjusted and an amount is subtracted from the gross income and passed to the tax authorities. If an individual owns shares and sells the shares making a profit, she is liable to pay income tax on the surplus. Companies making a profit have to pay income tax (but the name of the tax might be corporation or corporate tax); it is a tax on a company's income.

direct taxes a tax levied on income and wealth

Taxes on expenditure are referred to as indirect taxes. An **indirect tax** might be levied on a business that is responsible to the tax authorities to pay the tax but the business might pass on the tax to the consumer in the form of a higher price. Hence, the individual shares the burden of the tax and so is indirectly responsible for the tax.

indirect tax a tax levied on the sale of goods and services

Taxes on the sale of goods and services distort the market price and can lead to unintended consequences such as the industry shrinking.



We can further identify two types of tax on expenditure – a specific tax and an *ad valorem* tax. A **specific tax** is a set amount per unit of expenditure, for example, €0.75 per litre of petrol or €2.50 on a bottle of whisky. An **ad valorem tax** is expressed as a percentage, for example a 10 per cent tax or a 20 per cent tax. There is a difference in the way in which these types of taxes affect market outcomes. To analyse these outcomes we need to address a simple but subtle question: when the government levies a tax on a good, who bears the burden of the tax – in other words, who actually pays the tax? The people buying the good? The people selling the good? Or, if buyers and sellers share the tax burden, what determines how the burden is divided? Can the government simply legislate the division of the burden, or is the division determined by more fundamental forces in the economy? Economists use the term **tax incidence** to refer to the distribution of a tax burden. As we will see, some surprising lessons about tax incidence arise just by applying the tools of supply and demand.

specific tax a fixed rate tax levied on goods and services expressed as a sum per unit
ad valorem tax a tax levied as a percentage of the price of a good
tax incidence the manner in which the burden of a tax is shared among participants in a market

How Taxes on Sellers Affect Market Outcomes

We are going to analyse the market outcomes of a government imposing a specific tax and an *ad valorem* tax on sellers.

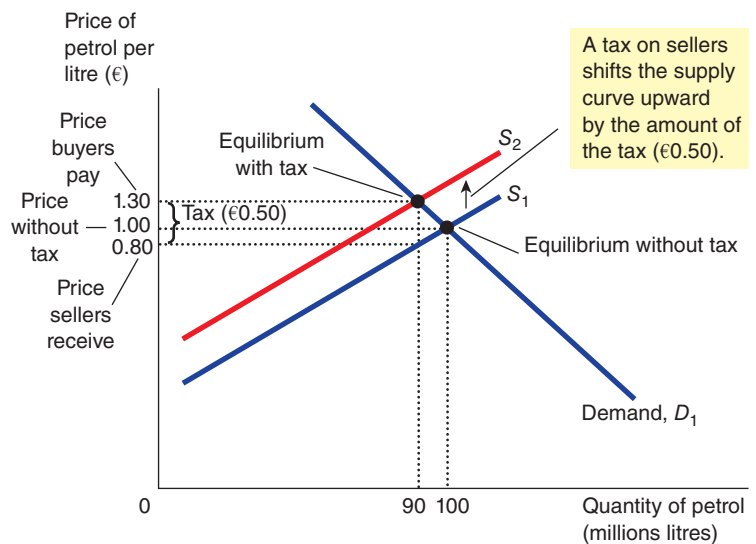
A Specific Tax Suppose the government imposes a tax on sellers of petrol of €0.50 for each litre of fuel they sell. What are the effects of this tax? Again, we apply our three steps.

Step One In this case, the immediate impact of the tax is on the sellers of petrol. The quantity of petrol demanded at any given price is the same; thus, the demand curve does not change. By contrast, the tax on sellers makes the petrol business less profitable at any given price – whatever the seller receives per litre, they now have to give €0.50 to the government. The seller is effectively facing an increase in the cost of production of €0.50 per litre and so this shifts the supply curve upwards.

FIGURE 8.4

A Specific Tax on Sellers

When a tax of €0.50 is levied on sellers, the supply curve shifts to the left by €0.50 at every price from S_1 to S_2 . The equilibrium quantity falls from 100 to 90 million litres. The price that buyers pay rises from €1.00 to €1.30 per litre. The price that sellers receive (after paying the tax) falls from €1.00 to €0.80. Even though the tax is levied on sellers, buyers and sellers share the burden of the tax.



Step Two Because the tax on sellers raises the cost of producing and selling petrol, it reduces the quantity supplied at every price. The supply curve shifts to the left (or, equivalently, upwards) and the shift is parallel to the original supply curve. The shift in the supply curve is a parallel one because regardless of the quantity supplied the seller has to pay the same amount per litre and so at every price the distance between the original and the new supply curve is the amount of the tax – €0.50.

For any market price of petrol, the effective price to sellers – the amount they get to keep after paying the tax – is €0.50 lower. For example, if the market price of petrol happened to be €2.00 per litre, the effective price received by sellers would be €1.50. Whatever the market price, sellers will supply a quantity of petrol as if the price were €0.50 lower than it is. Put differently, to induce sellers to supply any given quantity, the market price must now be €0.50 higher to compensate for the effect of the tax. Thus, as shown in Figure 8.4, the supply curve shifts *upward* from S_1 to S_2 by exactly the size of the tax (€0.50).

Step Three Having determined how the supply curve shifts, we can now compare the initial and the new equilibrium. Figure 8.4 shows that the equilibrium price of petrol rises from €1.00 to €1.30, and the

equilibrium quantity falls from 100 to 90 million litres. The tax reduces the size of the petrol market and buyers and sellers share the burden of the tax. Because the market price rises, buyers pay €0.30 more for each litre of petrol than they did before the tax was imposed. Sellers receive a higher price than they did without the tax, but the effective price (after paying the tax) falls from €1.00 to €0.80 per litre.

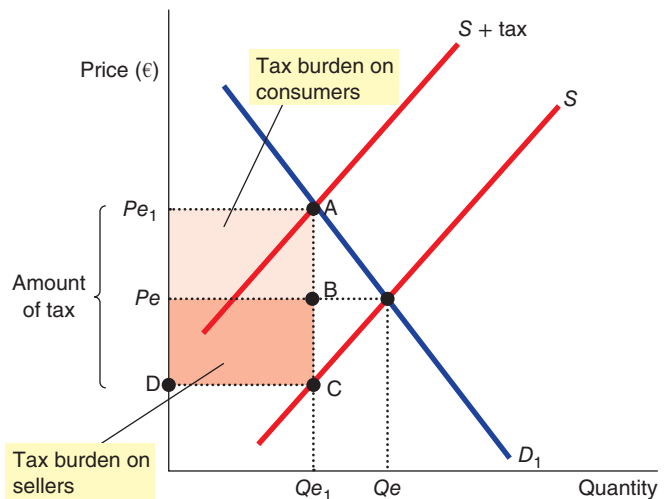
The total amount of tax paid by buyers and sellers can also be determined from Figure 8.4. Buyers pay an additional €0.30 per litre times the amount of petrol purchased (90 million litres) and so the total tax paid by buyers is €27 million. The burden of the tax on sellers is €0.20 per litre and they sell 90 million litres so sellers contribute €18 million to the tax authorities. The total tax revenue is the vertical distance between the two supply curves multiplied by the amount bought and sold. In this example, the total tax raised is $€0.50 \times 90 \text{ million} = €45 \text{ million}$.

This illustrates a general principle highlighted in Figure 8.5. The original equilibrium price before the tax is P_e and the original equilibrium quantity is Q_e . The levying of a tax (in this case a specific tax) shifts the supply curve to the left to $S + \text{tax}$. The new equilibrium price is P_{e_1} and the new equilibrium quantity is Q_{e_1} . The amount of the tax is the vertical distance between the two supply curves at the new equilibrium (AC). Buyers now pay a price of P_{e_1} compared to P_e and so pay $P_{e_1}P_e \times Q_{e_1}$ in tax, shown by the shaded rectangle $P_{e_1}ABP_e$. Sellers now receive D whereas they received P_e before the tax was levied. As a result the burden of the tax for sellers is the amount $DP_e \times Q_{e_1}$. The total amount paid by sellers, therefore, is given by the shaded area $PeBCD$. The total tax revenue due to the tax authorities is the area Pe_1ACD .

FIGURE 8.5

Determining the Incidence (Burden) of Taxation

A specific tax of AC per unit shifts the supply curve from S to $S + \text{tax}$. Consumers now pay a higher price given by P_{e_1} , and buy Q_{e_1} . The tax burden on the consumer is shown by the value of the shaded area $P_{e_1}ABP_e$. Sellers now receive the amount D for each unit sold compared to P_e before the tax was levied. The burden of the tax on sellers, therefore, is given by the value of the shaded area $PeBCD$. The total tax revenue raised and due to the authorities is the area Pe_1ACD .



Implications A tax on sellers places a wedge between the price that buyers pay and the price that sellers receive. The wedge between the buyers' price and the sellers' price is the same, and would be the same regardless of whether the tax is levied on buyers or sellers. In reality most governments levy taxes on sellers rather than on buyers, however. The wedge shifts the relative position of the supply and demand curves. In the new equilibrium, buyers and sellers share the burden of the tax.

An Ad Valorem Tax on Sellers Many readers of this text will be familiar with value added tax (VAT). VAT is a tax levied as a percentage. The basic principle that the buyer and seller both share the burden of the tax is the same as that for a specific tax but there is a subtle difference in the way the supply curve shifts because it is an *ad valorem* tax, which we will outline.

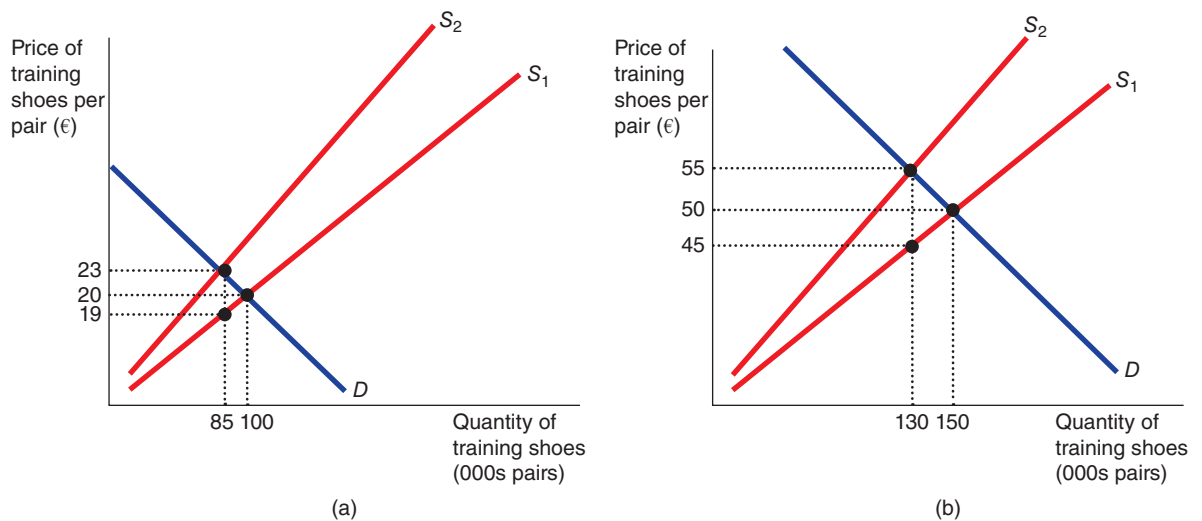
Imagine the market for training shoes where the government announces that it is going to impose a sales tax of 20 per cent. Again, we use our three-step method to analyse the effect.

Step One The impact of the tax is again on the sellers. The quantity of training shoes demanded at any given price is the same; thus, the demand curve does not change. The seller again faces an increase in the cost of production but this time the effective increase in cost varies at each price. If the tax was 20 per cent and training shoes cost €20 to produce, the seller would have to give €4 to the government in tax (20 per cent of €20); if training shoes cost €50 to produce, the seller would have to give €10 to the government and if training shoes cost €75 to produce the seller would have to give €15 to the government. The supply curve shifts to the left but it is not a parallel shift.

FIGURE 8.6**An Ad Valorem Tax on Sellers**

When a tax of 20 per cent is levied on sellers, the supply curve shifts to the left from S_1 to S_2 . At low prices, the amount of the tax paid is relatively low but 20 per cent of higher prices means the seller has to give more to the government. The shift in the supply curve is, therefore, not parallel. The market outcome will vary depending on the demand for trainers and the original market price. If market price were €20, as shown in panel (a), the equilibrium quantity falls from 100,000 to 85,000 pairs. The price that buyers pay rises from €20 to €23 per pair. The price that sellers receive (after paying the tax) falls from €20 to €19. Even though the tax is levied on sellers, buyers and sellers share the burden of the tax.

In panel (b) the equilibrium price before the tax is €50 per pair and the equilibrium quantity bought and sold is 150,000 pairs of trainers. The tax of 20 per cent means that the vertical distance between the two supply curves is now €10 which is how much the supplier has to give to the tax authorities for every pair sold. The buyer now faces a price of €55 per pair and the price the seller receives (after paying the tax) falls from €50 to €45.



Step Two The tax on sellers raises the cost of producing and selling trainers as for a specific tax but the amount that sellers have to give to the government is lower at low prices than at high prices because 20 per cent of a small amount is a different value than 20 per cent of a higher amount. The supply curve shifts to the left (or, equivalently, upwards) and the curve pivots upwards and to the left of the original supply curve as shown in Figure 8.6. At lower prices the seller has to pay a smaller amount per pair than at higher prices. The vertical distance between the supply curves at every price is 20 per cent of the price without the tax.

For any market price of training shoes, the effective price to sellers – the amount they get to keep after paying the tax – is 20 per cent lower. For example, if the market price of trainers happened to be €20 per pair, the effective price received by sellers would be €16. Whatever the market price, sellers will supply a quantity of trainers as if the price were 20 per cent lower than it is. Put differently, to induce sellers to supply any given quantity, the market price must now be 20 per cent higher to compensate for the effect of the tax. Thus, as shown in Figure 8.6, the supply curve shifts *upward* from S_1 to S_2 by 20 per cent at each price.

Step Three Having determined how the supply curve shifts, we can now compare the initial and the new equilibrium. In panel (a) the figure shows that the equilibrium price of training shoes rises from €20 to €23, and the equilibrium quantity falls from 100,000 pairs to 85,000 pairs. The tax reduces the size of the training shoe market and buyers and sellers share the burden of the tax. Because the market price rises, buyers pay €3 more for each pair of trainers than they did before the tax was imposed. Sellers receive a higher price than they did without the tax, but the effective price (after paying the tax) falls from €20 to €19 per pair.

In panel (b) the initial equilibrium price before the tax is €50 and the amount of training shoes bought and sold is 150,000 pairs. A 20 per cent tax will mean that sellers have to pay the tax authorities €10 on each pair sold. The figure shows the equilibrium price has risen to €55 so buyers have to pay €5 more per pair of trainers and the price that sellers actually receive falls from €50 before the tax to €45 afterwards. In this example, the burden of the tax is shared equally between the buyer and seller.

The Algebra of a Specific Tax Assume that the demand function is given by the equation:

$$P_d = 30 - 1.5q$$

and the supply equation by:

$$P_s = 6 + 0.5q$$

A specific tax levied on the seller of t would make the supply function:

$$P_s = 6 + 0.5q + t$$

In equilibrium, therefore:

$$\begin{aligned} 6 + 0.5q + t &= 30 - 1.5q \\ 0.5q + 1.5q &= 30 - 6 - t \\ 2q &= 24 - t \\ q &= 12 - 0.5t \end{aligned}$$

If no tax was levied then the quantity would be 12. If the tax t was levied at a rate of €6 per unit then the quantity would be 9 and if the tax was levied at a rate of €8 per unit, $q = 8$.

We can also substitute $q = 12 - 0.5t$ into the demand equation and determine the effect on price of the tax. The demand equation would thus be:

$$P_d = 30 - 1.5(12 - 0.5t)$$

If $t = 5$ then the price would be:

$$\begin{aligned} P_d &= 30 - 1.5(12 - 0.5(5)) \\ P_d &= 30 - 14.25 \\ P_d &= 15.75 \end{aligned}$$

If $t = 8$ then the price would be:

$$\begin{aligned} P_d &= 30 - 1.5(12 - 4) \\ P_d &= 30 - 12 \\ P_d &= 18 \end{aligned}$$

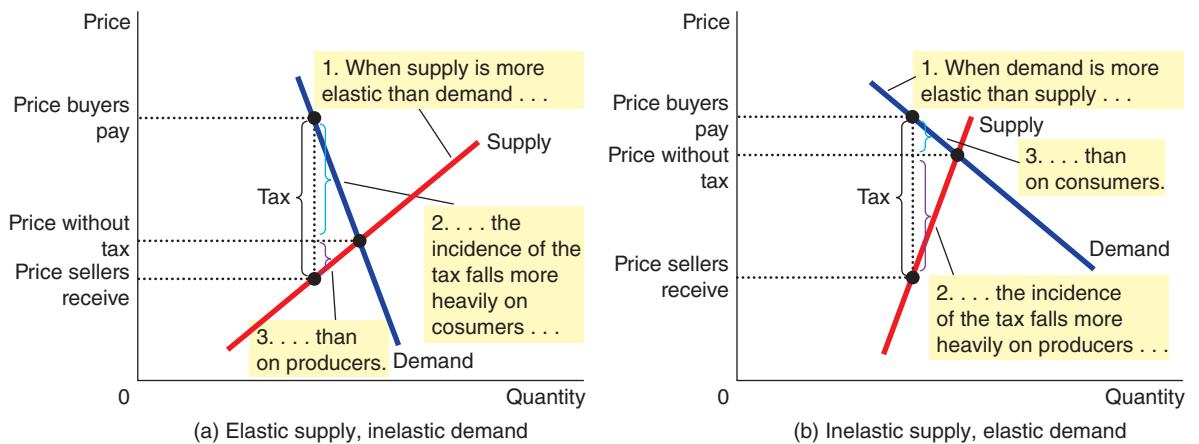
Elasticity and Tax Incidence

When a good is taxed, buyers and sellers of the good share the burden of the tax. But how exactly is the tax burden divided? Only rarely will it be shared equally. To see how the burden is divided, consider the impact of taxation in the two markets in Figure 8.7. In both cases, the figure shows the initial demand

curve, the initial supply curve, and a tax that drives a wedge between the amount paid by buyers and the amount received by sellers. (Not drawn in either panel of the figure is the new supply or demand curve – which curve shifts depends on whether the tax is levied on buyers or sellers. As we have seen, this is irrelevant for the incidence of the tax.)

FIGURE 8.7**How the Burden of a Tax Is Divided**

In panel (a), the supply curve is elastic and the demand curve is inelastic. In this case, the price received by sellers falls only slightly, while the price paid by buyers rises substantially. Thus, buyers bear most of the burden of the tax. In panel (b) the supply curve is inelastic and the demand curve is elastic. In this case, the price received by sellers falls substantially, while the price paid by buyers rises only slightly. Thus, sellers bear most of the burden of the tax.



Panel (a) of Figure 8.7 shows a tax levied in a market with very elastic supply and relatively inelastic demand. That is, sellers are very responsive to changes in the price of the good (so the supply curve is relatively flat), whereas buyers are not very responsive (so the demand curve is relatively steep). When a tax is imposed on a market with these elasticities, the price received by sellers does not fall much, so sellers bear only a small burden. By contrast, the price paid by buyers rises substantially, indicating that buyers bear most of the burden of the tax. Our analysis of elasticity in Chapter 4 should make this something that is not at all surprising. If the price elasticity of demand is low (a steep curve) then demand will fall proportionately less in response to a rise in price – buyers are not very price sensitive. The seller can shift the burden of the tax onto the buyer safe in the knowledge that demand will only fall by a relatively small amount.

Panel (b) of Figure 8.7 shows a tax in a market with relatively inelastic supply and very elastic demand. In this case, sellers are not very responsive to changes in the price (so the supply curve is steeper), while buyers are very responsive (so the demand curve is flatter). The figure shows that when a tax is imposed, the price paid by buyers does not rise much, while the price received by sellers falls substantially. Thus, sellers bear most of the burden of the tax. In this case, sellers know that if they try to pass on the tax to buyers that demand will fall by a relatively large amount.

The two panels of Figure 8.7 show a general lesson about how the burden of a tax is divided: a tax burden falls more heavily on the side of the market that is less elastic. Why is this true? In essence, the elasticity measures the willingness of buyers or sellers to leave the market when conditions become unfavourable. A small elasticity of demand means that buyers do not have good alternatives to consuming this particular good. A small elasticity of supply means that sellers do not have good alternatives to producing this particular good. When the good is taxed, the side of the market with fewer good alternatives cannot easily leave the market and must, therefore, bear more of the burden of the tax.

SELF TEST In a supply and demand diagram, show how a specific tax on car buyers of €1,000 per car affects the quantity of cars sold and the price of cars. In another diagram, show how a tax on car sellers of €1,000 per car affects the quantity of cars sold and the price of cars. In both of your diagrams, show the change in the price paid by car buyers and the change in price received by car sellers.

In some countries governments levy both a specific tax (such as duties) and a sales tax such as VAT on products. Show the market outcome for cars if the government imposes a specific tax of €500 per car and VAT at a rate of 15 per cent.

SUBSIDIES

A subsidy is the opposite of a tax. A **subsidy** is a payment to buyers and sellers to supplement income or reduce costs of production to provide an advantage to the recipient of the subsidy. Subsidies are levied when governments want to encourage the consumption of a 'good' which they deem is currently under-produced. Taxes, on the other hand, may be levied on a 'bad' which the government believes is over-consumed. Subsidies are generally given to sellers and have the effect of reducing the cost of production as opposed to a tax which increases the cost of production. Subsidies exist in a variety of different areas including education, transport, agriculture, regional development, housing and employment.

subsidy payment to buyers and sellers to supplement income or lower costs and which thus encourages consumption or provides an advantage to the recipient

How Subsidies Affect Market Outcomes

Most European countries provide subsidies for transport systems and the common agricultural policy oversees subsidies to farmers. In Switzerland some €2.5 billion is spent on subsidies for rail transport, in Germany the figure is nearer to €9 billion whilst in the UK subsidies account for around €3 billion and in France €6.8 billion.

Figure 8.8 shows how a subsidy works using the rail system as an example and utilizing our three steps. In the absence of a subsidy the equilibrium number of journeys bought and sold is Q_e and the equilibrium price for the average train ticket for each journey is €75.

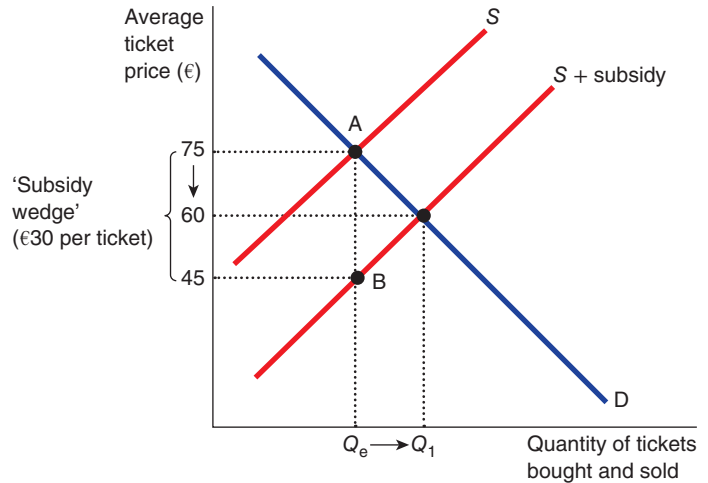
Step One If the government gives a subsidy of €30 per (average) ticket to train operators, it is the supply curve for journeys which is affected; the demand curve is not affected because the number of train journeys demanded at each price stays the same. The subsidy to train operators reduces the cost of providing a train journey by €30 and so the supply curve will shift.

Step Two Because the subsidy reduces the cost to the train operators, the supply curve shifts to the right by the amount of the subsidy from S to $S + \text{subsidy}$. The amount of the subsidy per ticket is shown by the vertical distance between the two supply curves, AB. If the cost of providing a train journey was an average of €75 and the subsidy was €30 the supply curve would shift so that train operators would now supply train journeys at an effective cost of €30 below the previous cost. They would now be willing to supply more train journeys at every price.

Step Three Comparing the initial and the new equilibrium we can see that the equilibrium price of each train journey is now lower at €60 and the equilibrium number of journeys travelled increases to Q_1 .

FIGURE 8.8**A Subsidy on Rail Transport**

When a subsidy of €30 per ticket is given to sellers, the supply curve shifts to the right from S to $S + \text{subsidy}$. The equilibrium quantity rises from Q_e to Q_1 journeys per year. The price that buyers pay for a journey falls from €75 to €60. The subsidy results in lower prices for passengers and an increased number of journeys available. Even though the subsidy is given to sellers, buyers and sellers share the benefits of the subsidy.



Passengers and train operators both benefit from the subsidy, as passengers can obtain train tickets at a lower price than before the subsidy and have more journeys available, and sellers receive more revenue than they did before the subsidy allowing them the potential to invest in the service they provide. The precise division of the benefits between buyers and sellers will depend on the relative elasticities of demand and supply.

The total cost of providing the subsidy is the amount of the subsidy (€30 per ticket in this case) multiplied by the number of tickets bought and sold (Q_e) shown by the area 75, A, B, 45.

Implications There is a considerable debate surrounding the value of subsidies. We have seen from the example how price and quantity can be affected following the imposition of a subsidy. In the case of transport, it may have the effect of altering the incentives for people to travel on the train rather than on the roads and so have the benefit of reducing congestion on the roads as well as reducing possible pollution that is associated with road use. There are also costs associated with subsidies; for one thing someone has to finance the subsidy and it is often the taxpayer. Subsidies may also encourage firms to overproduce, which has a wider effect on the market. Subsidies on commodities such as cotton, bananas and sugar distort the workings of the market and change global comparative advantage. Overproduction leads to excess supply on world markets and drives down prices as well as diverting trade to rich countries who can support producers through subsidies at the expense of poor countries whose producers cannot compete because prices are lower than the free market price.

CONCLUSION

The economy is governed by two kinds of laws: the laws of supply and demand and the laws enacted by governments. In this chapter we have begun to see how these laws interact. Price controls, taxes and subsidies are common in various markets in the economy, and their effects are frequently debated in the press and among policymakers. Even a little bit of economic knowledge can go a long way towards understanding and evaluating these policies.

In the next chapter we shall examine the effects of taxation and subsidies more fully, and we shall consider a broader range of policies than we considered here. Yet the basic lessons of this chapter will not change: when analysing government policies, supply and demand are the first and most useful tools of analysis.

IN THE NEWS



Price Floors

Despite the criticism from some about the costs of governments interfering in markets to produce 'desired outcomes' there continues to be moves to do so. One more recent example in the UK was an announcement by the government to introduce a carbon price floor.

A Carbon Price Floor

In the 2011 Budget speech, the UK government announced plans for a carbon price floor which would come into operation in April 2013. The continued concern over carbon emissions and the effect increased carbon emissions might be having on the long-term climate means that the government is looking at measures to try to encourage firms to reduce carbon emissions. Firms affected by the carbon price floor include those generating fossil-fuel based electricity, heat and power operators and electricity firms. Critics of the move say that many manufacturing firms will also be affected and that the policy will reduce competitiveness.

The carbon price floor will be set at £16 per tonne and will then rise each year by £2 per tonne to reach

£30 per tonne by 2020. By comparison, the rate set in the rest of Europe is £5.60 per tonne but might rise to a maximum of £9.70.

By setting such a price floor, the UK government is hoping that firms will look at ways of investing in low-carbon forms of energy. Key to its success, however, is the level it is set at. If the floor is set too low then firms might absorb any increased costs that result and see the increased price as a tax and not look for other low-carbon options. If it is set too high, however, firms will complain that in the difficult economic environment in which they are operating this will merely make them less competitive, especially if the levels set in the rest of Europe are much lower. A higher floor might be more of an incentive to invest in low-carbon energy and make such

an investment more commercially viable.

Business leaders have voiced concerns on the level of the floor and the 'escalator' which will see the floor increase up to 2020 and these were raised even further when the UK government announced that Northern Ireland would be exempt from the scheme in the Autumn Budget Statement in December 2012. There were suggestions that the reason for the exemptions was because firms in the Republic of Ireland would not be subject to any such floor and this would mean that Northern Ireland firms would be at a competitive disadvantage.

Questions

- 1 For the carbon price floor to be effective, what market conditions must exist?**
- 2 Using supply and demand diagrams, explain how a carbon price floor is meant to reduce carbon emissions.**
- 3 Consider the factors which might contribute to the success of the carbon price floor in achieving its goal.**
- 4 What might be the possible costs of imposing a carbon price floor?**
- 5 What does the decision of the UK government to exempt Northern Ireland from the scheme tell you about whether business leaders are right to be concerned that the scheme will reduce competitiveness of UK firms?**

The carbon price floor is aimed at encouraging firms using fossil fuels to invest in cleaner technologies.



SUMMARY

- A price ceiling is a legal maximum on the price of a good or service. An example is rent control. If the price ceiling is below the equilibrium price, the quantity demanded exceeds the quantity supplied. Because of the resulting shortage, sellers must in some way ration the good or service among buyers.
- A price floor is a legal minimum on the price of a good or service. If the price floor is above the equilibrium price, the quantity supplied exceeds the quantity demanded. Because of the resulting surplus, buyers' demands for the good or service must in some way be rationed among sellers.
- When the government levies a tax on a good, the equilibrium quantity of the good falls. That is, a tax on a market shrinks the size of the market.
- A tax on a good places a wedge between the price paid by buyers and the price received by sellers. When the market moves to the new equilibrium, buyers pay more for the good and sellers receive less for it. In this sense, buyers and sellers share the tax burden. The incidence of a tax (that is, the division of the tax burden) does not depend on whether the tax is levied on buyers or sellers.
- A subsidy given to sellers lowers the cost of production and encourages firms to expand output. Buyers benefit from lower prices.
- The incidence of a tax or subsidy depends on the price elasticities of supply and demand. The burden tends to fall on the side of the market that is less elastic because that side of the market can respond less/more easily to the tax/ subsidy by changing the quantity bought or sold.

QUESTIONS FOR REVIEW

- 1 Give an example of a price ceiling and an example of a price floor.
- 2 Which causes a shortage of a good – a price ceiling or a price floor? Which causes a surplus?
- 3 Under what circumstances is a price ceiling and a price floor referred to as binding?
- 4 What potential costs and benefits might a government have to consider in deciding whether to impose a price floor or a price ceiling?
- 5 What mechanisms allocate resources when the price of a good is not allowed to bring supply and demand into equilibrium?
- 6 Explain why economists usually oppose rent controls.
- 7 How does a tax imposed on a good with a high price elasticity of demand affect the market equilibrium? Who bears most of the burden of the tax in this instance?
- 8 How does a subsidy on a good affect the price paid by buyers, the price received by sellers and the quantity bought and sold?
- 9 What determines how the burden of a tax or a subsidy is divided between buyers and sellers? Why?
- 10 How might an economist decide whether the benefits of a subsidy outweigh the cost?

PROBLEMS AND APPLICATIONS

- 1 Lovers of opera persuade the government to impose a price ceiling of €50 per ticket at the country's national opera house. Does this policy get more or fewer people to attend? What does the market outcome depend on?
- 2 The government has decided that the free market price of tobacco is too low.
 - a. Suppose the government imposes a binding price floor in the tobacco market. Use a supply and demand diagram to show the effect of this policy on the price of tobacco and the quantity of tobacco sold. Is there a shortage or surplus of tobacco? What does the market outcome depend on?
 - b. Tobacco producers complain that the price floor has reduced their total revenue. Is this possible? Explain.
 - c. In response to producers' complaints, the government agrees to purchase all of the surplus tobacco

at the price floor. Compared to the basic price floor, who benefits from this new policy? Who loses?

- 3 A recent study found that the demand and supply schedules for frisbees are as follows:

| Price per frisbee | Quantity demanded | Quantity supplied |
|-------------------|--------------------|---------------------|
| €11 | 1 million frisbees | 15 million frisbees |
| 10 | 2 | 12 |
| 9 | 4 | 9 |
| 8 | 6 | 6 |
| 7 | 8 | 3 |
| 6 | 10 | 1 |

What are the equilibrium price and quantity of frisbees?

- a. Frisbee manufacturers persuade the government that frisbee production improves scientists' understanding of aerodynamics and thus is important for

- national security. The government decides to impose a price floor €2 above the equilibrium price. What is the new market price? How many frisbees are sold?
- b. Irate university students march on the government and demand a reduction in the price of frisbees. The government decides to repeal the price floor and impose a price ceiling €1 below the former price floor. What is the new market price? How many frisbees are sold?
- 4 Suppose that, in the absence of any tax whatsoever, the equilibrium price of a flat is €150,000. Now suppose that the government requires buyers in the housing market to pay a tax of €5,000 on every flat purchased.
- a. Draw a supply and demand diagram of the market for flats without the tax. Show the price paid by consumers, the price received by producers and the quantity of flats bought and sold. What is the difference between the price paid by consumers and the price received by producers?
- b. Now draw a supply and demand diagram for the market for flats with the tax. Show the price paid by consumers, the price received by producers and the quantity of flats bought and sold. What is the difference between the price paid by consumers and the price received by producers? Has the quantity of flats bought and sold increased or decreased?
- 5 A country's finance minister wants to raise tax revenue to help reduce government borrowing. Advise the minister what type of goods s/he should consider levying a tax on to accomplish the minister's goal.
- 6 A government wants to reduce the incidence of anti-social behaviour which can result from excess alcohol consumption by young people aged between 18 and 25. Which do you think leads to a fairer market outcome: a specific tax on every unit of alcohol sold or an *ad valorem* tax on the price of alcohol equal to 25 per cent of the price per unit? Use diagrams to help support your answer to the question.
- 7 If the government places a €1,000 tax on luxury cars, will the price paid by consumers rise by more than €1,000, less than €1,000 or exactly €1,000. Explain.
- 8 The government decides to reduce air pollution by reducing the use of petrol. It imposes €0.50 tax for each litre of petrol sold.
- a. Should it impose this tax on petrol companies or motorists? Explain carefully, using a supply and demand diagram.
- b. If the demand for petrol were more elastic, would this tax be more effective or less effective in reducing the quantity of petrol consumed? Explain with both words and a diagram.
- c. Are consumers of petrol helped or hurt by this tax? Why?
- d. Are workers in the oil industry helped or hurt by this tax? Why?
- 9 Unemployment amongst young people across a number of European countries is very high. Imagine that European Union ministers proposed a youth employment subsidy designed to help increase the number of young people between the ages of 18 and 29 into work, by making it easier for employers to take on extra workers. Assume that estimates of the cost of the subsidy would be €2,500 per job and would create up to 175,000 new jobs. Use diagrams to show how the subsidy is designed to work and what its success might depend upon.
- 10 There is a debate in a country about the higher education system. Some political parties want to promote the value of university education and give as many young people the chance to attend university as possible. This party believe that the state should provide grants for young people to help them go to university. An opposing party also agrees that a university education is a good thing but believes that the benefits that a university education provides should be paid for by those who go to university and not by the state. They advocate students paying the tuition fees for their university at the start of each academic year.

Present an argument for and against both policies, and state which you think is the most efficient and equitable market outcome. Use diagrams to help support your answer where appropriate.

PART 4

THE ECONOMICS OF THE PUBLIC SECTOR

9

THE TAX SYSTEM AND THE COSTS OF TAXATION

Taxes have been around a long time. In the Bible, for example, we can read how Jesus's parents had to return to Nazareth to be taxed and how, later, Jesus converted a prominent tax collector to become one of his disciples. In the Koran, there are references to *jizya*, which can be translated as the word 'tax'. In the Ottoman Empire *jizya* was a per-capita tax levied on non-Muslim, able-bodied resident males of military age. Since taxes, by definition, are a means of legally extracting money from individuals or organizations, it is not surprising that they have often been a source of heated political debate throughout history.

In Chapter 8 we looked at taxes and subsidies and the incidence of taxation. In this chapter we are going to explore some of the theory behind the design of tax systems and consider the fundamental principles of taxation. Most people agree that taxes should impose as small a cost on society as possible and that the burden of taxes should be distributed fairly. That is, the tax system should be both efficient and equitable. We will then use the tools developed in Chapter 7 to analyse the welfare effects of taxation.

TAXES AND EFFICIENCY

Most governments levy taxes for two main reasons. One is to raise revenue to help pay for the various services that government provides. The second reason is to influence behaviour and achieve market outcomes that are deemed desirable. For example, governments might choose to levy high taxes on tobacco and alcohol to curb consumption; taxes on certain types of energy production could be reduced or even made zero to try and encourage a switch from consumption of energy which is deemed damaging to the environment (or which will ultimately become scarcer and so subject to rising prices) to energy generation and consumption which is renewable and 'green'. There are many ways to achieve these different aims but in designing a tax system, policymakers have two objectives: efficiency and equity.

One tax system is more efficient than another if it raises the same amount of revenue at a smaller cost to taxpayers and the government. What are the costs of taxes to taxpayers? The most obvious cost is the tax payment itself. This transfer of money from the taxpayer to the government is an inevitable feature of any tax system. Yet taxes also impose two other costs, which well-designed tax policy tries to avoid or, at least, minimize. Taxes and subsidies affect consumer and producer behaviour and produce different

market outcomes compared to free market outcomes. We can attempt to measure the welfare effects of these different market outcomes by looking at the changes to consumer and producer surplus that result. Any reduction in total surplus in the market outcome when taxes are levied compared to a free market outcome is called the **deadweight loss**.

deadweight loss the fall in total surplus that results from a market distortion, such as a tax

The second cost we need to be aware of is the administrative burdens that taxpayers bear as they comply with the tax laws.

An efficient tax system is one that imposes small deadweight losses and small administrative burdens.

THE DEADWEIGHT LOSS OF TAXATION

One of the *Ten Principles of Economics* is that people respond to incentives, and this includes incentives provided by the tax system. Taxes have the effect of changing people's behaviour. If the government taxes tea, some people will drink less tea and drink more coffee. If the government taxes housing, some people will change behaviour and look to live in smaller houses and spend more of their income on other things. If the government taxes labour earnings, some people will not see additional work as having the same reward and may decide to work less and enjoy more leisure.

The effects of taxes on welfare might at first seem obvious. If the government imposes taxes in order to raise revenue, that revenue must come out of someone's pocket. When a good is taxed, both buyers and sellers are worse off: a tax raises the price buyers pay and lowers the price sellers receive. Yet to understand fully how taxes affect economic well-being, we must compare the reduced welfare of buyers and sellers to the amount of revenue the government raises.

Because taxes have the effect of changing behaviour and distorting incentives, they entail deadweight losses which reduce economic well-being of taxpayers in excess of the amount of revenue raised by the government. The deadweight loss is the inefficiency that a tax creates as people allocate resources according to the tax incentive rather than the true costs and benefits of the goods and services that they buy and sell.

In Chapter 8 we looked at the effect of a tax levied on a seller. The same result would be gained, however, if we had analysed the effect of a tax on a buyer. When a tax is levied on buyers, the demand curve shifts downwards by the size of the tax; when it is levied on sellers, the supply curve shifts upwards by that amount. In either case, when the tax is imposed, the price paid by buyers rises, and the price received by sellers falls. In the end, buyers and sellers share the burden of the tax, regardless of how it is levied.

Figure 9.1 shows these effects. To simplify our discussion, this figure does not show a shift in either the supply or demand curve, although one curve must shift. Which curve shifts depends on whether the tax is levied on sellers (the supply curve shifts) or buyers (the demand curve shifts). In this chapter, we can simplify the graphs by not bothering to show the shift. The key result for our purposes here is that the tax places a wedge between the price buyers pay and the price sellers receive. Because of this tax wedge, the quantity sold falls below the level that would be sold without a tax. In other words, a tax on a good causes the size of the market for the good to shrink. These results should be familiar from Chapter 8.

How a Tax Affects Market Participants

Now let's use the tools of welfare economics to measure the gains and losses from a tax on a good. To do this, we must take into account how the tax affects buyers, sellers and the government. We know that the benefit to buyers and sellers can be measured by looking at the changes in consumer and producer surplus. What about the third interested party, the government? If T is the size of the tax and Q is the quantity of the good sold, then the government gets total tax revenue of $T \times Q$. It can use this tax revenue to provide services, such as roads, police and education, or to help the needy. Therefore, to analyse how taxes affect economic well-being, we use tax revenue to measure the government's benefit from the tax. Keep in mind, however, that this benefit actually accrues not to government but to those on whom the revenue is spent.

FIGURE 9.1**The Effects of a Tax**

A tax on a good places a wedge between the price that buyers pay and the price that sellers receive. The quantity of the good sold falls.

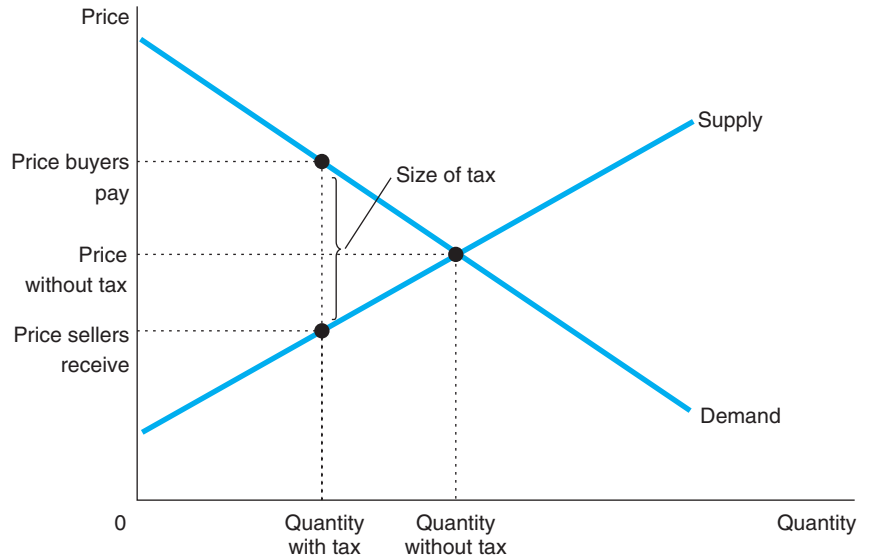
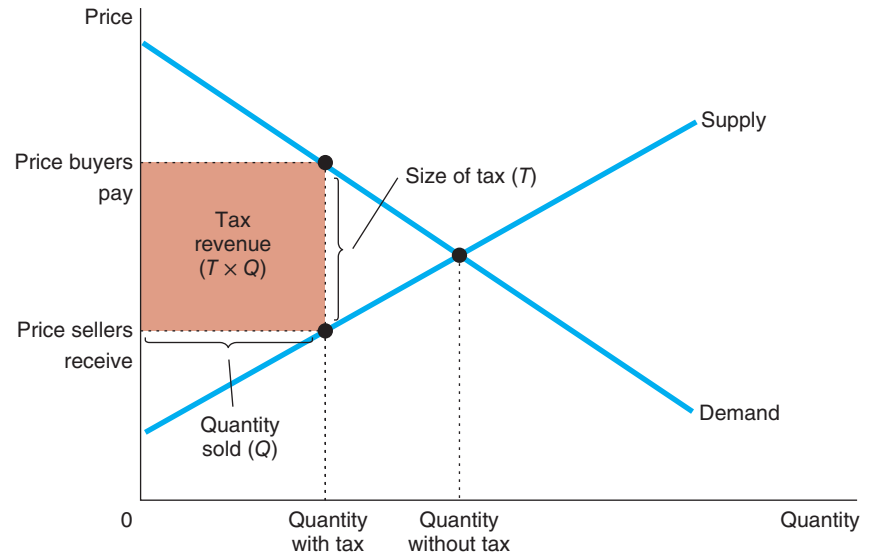


Figure 9.2 shows that the government's tax revenue is represented by the rectangle between the supply and demand curves. The height of this rectangle is the size of the tax, T , and the width of the rectangle is the quantity of the good sold, Q . Because a rectangle's area is its height times its width, this rectangle's area is $T \times Q$, which equals the tax revenue.

FIGURE 9.2**Tax Revenue**

The tax revenue that the government collects equals $T \times Q$, the size of the tax T times the quantity sold Q . Thus, tax revenue equals the area of the rectangle between the supply and demand curves.



Welfare Without a Tax To see how a tax affects welfare, we begin by considering welfare before the government has imposed a tax. Figure 9.3 shows the supply and demand diagram and marks the key areas with the letters A to F.

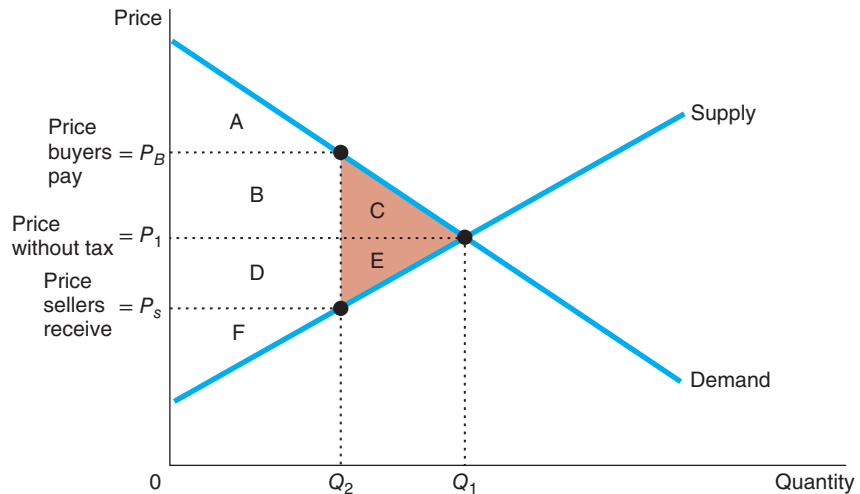
Without a tax, the price and quantity are found at the intersection of the supply and demand curves. The price is P_1 , and the quantity sold is Q_1 . Because the demand curve reflects buyers' willingness to pay, consumer surplus is the area between the demand curve and the price, $A + B + C$. Similarly, because the

supply curve reflects sellers' costs, producer surplus is the area between the supply curve and the price, $D + E + F$. In this case, because there is no tax, tax revenue equals zero.

Total surplus – the sum of consumer and producer surplus – equals the area $A + B + C + D + E + F$. In other words, total surplus is the area between the supply and demand curves up to the equilibrium quantity. The first column of the table in Figure 9.3 summarizes these conclusions.

FIGURE 9.3**How a Tax Affects Welfare**

A tax on a good reduces consumer surplus (by the area $B + C$) and producer surplus (by the area $D + E$). Because the fall in producer and consumer surplus exceeds tax revenue (area $B + D$), the tax is said to impose a deadweight loss (area $C + E$). The area $C + E$ shows the fall in total surplus and is the deadweight loss of the tax.



| | Without tax | With tax | Change |
|------------------|-------------------------|-----------------|------------|
| Consumer surplus | $A + B + C$ | A | $-(B + C)$ |
| Producer surplus | $D + E + F$ | F | $-(D + E)$ |
| Tax revenue | None | $B + D$ | $+(B + D)$ |
| Total surplus | $A + B + C + D + E + F$ | $A + B + D + F$ | $-(C + E)$ |

Welfare With a Tax Now consider welfare after the tax is imposed. The price paid by buyers rises from P_1 to P_B , so consumer surplus now equals only area A (the area below the demand curve and above the buyer's price). The price received by sellers falls from P_1 to P_S , so producer surplus now equals only area F (the area above the supply curve and below the seller's price). The quantity sold falls from Q_1 to Q_2 , and the government collects tax revenue equal to the area $B + D$.

To compute total surplus with the tax, we add consumer surplus producer surplus and tax revenue. Thus, we find that total surplus is area $A + B + D + F$. The second column of the table provides a summary.

Changes in Welfare We can now see the effects of the tax by comparing welfare before and after the tax is imposed. The third column in the table in Figure 9.3 shows the changes. The tax causes consumer surplus to fall by the area $B + C$ and producer surplus to fall by the area $D + E$. Tax revenue rises by the area $B + D$. Not surprisingly, the tax makes buyers and sellers worse off and the government better off.

The change in total welfare includes the change in consumer surplus (which is negative), the change in producer surplus (which is also negative), and the change in tax revenue (which is positive). When we add these three pieces together, we find that total surplus in the market falls by the area $C + E$. Thus, the losses to buyers and sellers from a tax exceed the revenue raised by the government. The fall in total surplus that results when a tax (or some other policy) distorts a market outcome is the deadweight loss. The area $C + E$ measures the size of the deadweight loss.

To understand why taxes impose deadweight losses, recall one of the *Ten Principles of Economics*: people respond to incentives. We have assumed that markets normally allocate scarce resources efficiently. That is, the equilibrium of supply and demand maximizes the total surplus of buyers and sellers in

a market. When a tax raises the price to buyers and lowers the price to sellers, however, it gives buyers an incentive to consume less and sellers an incentive to produce less than they otherwise would. As buyers and sellers respond to these incentives, the size of the market shrinks below its optimum. Thus, because taxes distort incentives, they cause markets to allocate resources inefficiently.

Deadweight Losses and the Gains from Trade

To gain some intuition for why taxes result in deadweight losses, consider an example. Imagine that Carsten cleans Annika's house each week for €100. The opportunity cost of Carsten's time is €80, and the value of a clean house to Annika is €120. Thus, Carsten and Annika each receive a €20 benefit from their deal. The total surplus of €40 measures the gains from trade in this particular transaction.

Now suppose that the government levies a €50 tax on the providers of cleaning services. There is now no price that Annika can pay Carsten that will leave both of them better off after paying the tax. The most Annika would be willing to pay is €120, but then Carsten would be left with only €70 after paying the tax, which is less than his €80 opportunity cost. Conversely, for Carsten to receive his opportunity cost of €80, Annika would need to pay €130, which is above the €120 value she places on a clean house. As a result, Annika and Carsten cancel their arrangement. Carsten goes without the income, and Annika lives in a dirtier house.

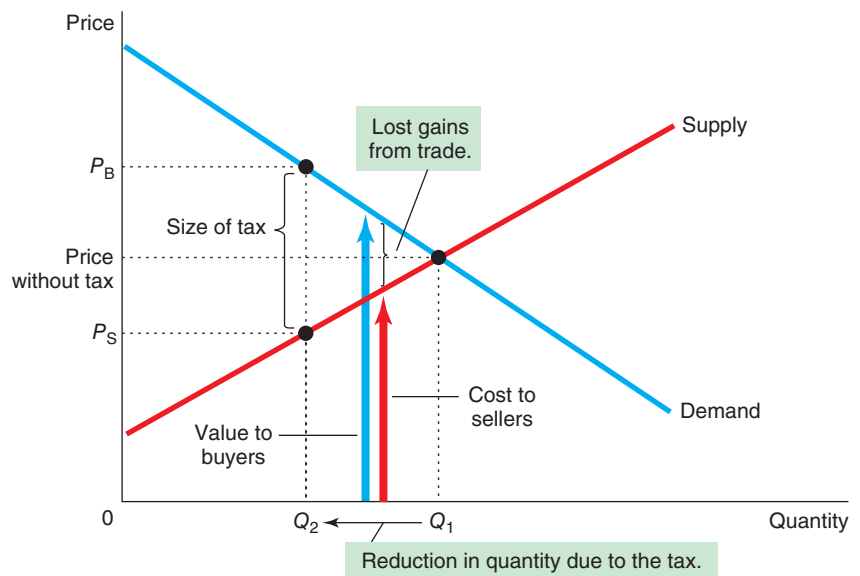
The tax has made Carsten and Annika worse off by a total of €40, because they have lost this amount of surplus. At the same time, the government collects no revenue from Carsten and Annika because they decide to cancel their arrangement. The €40 is pure deadweight loss: it is a loss to buyers and sellers in a market not offset by an increase in government revenue. From this example, we can see the ultimate source of deadweight losses: taxes cause deadweight losses because they prevent buyers and sellers from realizing some of the gains from trade.

The area of the triangle between the supply and demand curves (area C + E in Figure 9.3) measures these losses. This loss can be seen most easily in Figure 9.4 by recalling that the demand curve reflects the value of the good or the benefit to consumers and that the supply curve reflects the costs of producers. When the tax raises the price to buyers to P_B and lowers the price to sellers to P_S , the marginal buyers and sellers leave the market, so the quantity sold falls from Q_1 to Q_2 . Yet, as the figure shows, the value of the good to these buyers still exceeds the cost to these sellers. As in our example with Carsten and Annika, the gains from trade – the difference between buyers' value and sellers' cost – is less than the tax. Thus, these trades do not get made once the tax is imposed. The deadweight loss is the surplus lost because the tax discourages these mutually advantageous trades.

FIGURE 9.4

The Deadweight Loss

When the government imposes a tax on a good, the quantity sold falls from Q_1 to Q_2 . As a result, some of the potential gains from trade among buyers and sellers do not get realized. These lost gains from trade create the deadweight loss.



SELF TEST Draw the supply and demand curve for parking on business premises in city centres. If the government imposes a tax on parking spaces, show what happens to the quantity sold, the price paid by buyers and the price paid by sellers. In your diagram, show the deadweight loss from the tax. Explain the meaning of the deadweight loss.

Why might the government have imposed this tax? Do you think this was because of efficiency or equity reasons?

The Determinants of the Deadweight Loss

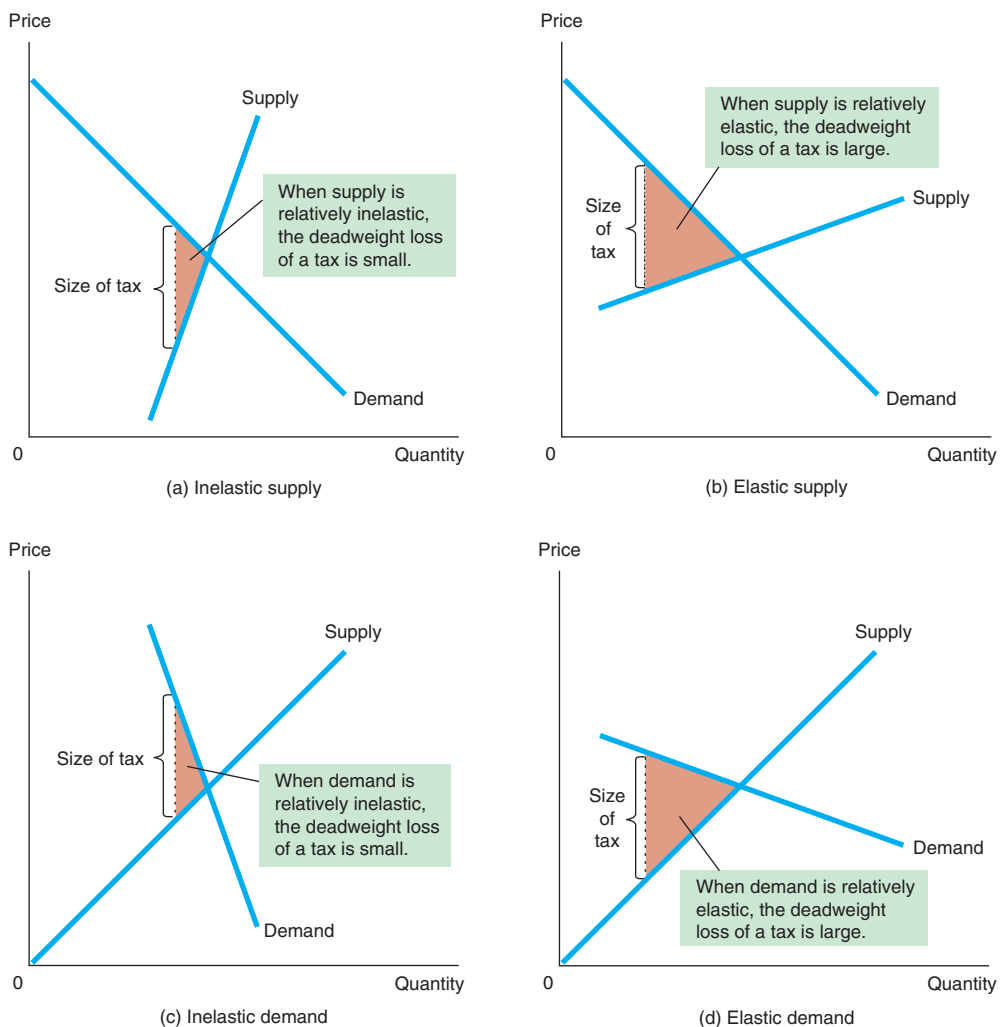
What determines whether the deadweight loss from a tax is large or small? The answer is the price elasticities of supply and demand.

Let's consider first how the elasticity of supply affects the size of the deadweight loss. In the top two panels of Figure 9.5, the demand curve and the size of the tax are the same. The only difference in these figures is the elasticity of the supply curve. In panel (a), the supply curve is relatively inelastic: quantity

FIGURE 9.5

Tax Distortions and Elasticities

In panels (a) and (b) the demand curve and the size of the tax are the same, but the price elasticity of supply is different. Notice that the more elastic the supply curve, the larger the deadweight loss of the tax. In panels (c) and (d) the supply curve and the size of the tax are the same, but the price elasticity of demand is different. Notice that the more elastic the demand curve, the larger the deadweight loss of the tax.



supplied responds only slightly to changes in the price. In panel (b), the supply curve is relatively elastic: quantity supplied responds substantially to changes in the price. Notice that the deadweight loss, the area of the triangle between the supply and demand curves, is larger when the supply curve is more elastic.

Similarly, the bottom two panels of Figure 9.5 show how the elasticity of demand affects the size of the deadweight loss. Here the supply curve and the size of the tax are held constant. In panel (c) the demand curve is relatively inelastic, and the deadweight loss is small. In panel (d) the demand curve is more elastic, and the deadweight loss from the tax is larger.

The lesson from Figure 9.5 is easy to explain. A tax has a deadweight loss because it induces buyers and sellers to change their behaviour. The tax raises the price paid by buyers, so they consume less. At the same time, the tax lowers the price received by sellers, so they produce less. Because of these changes in behaviour, the size of the market shrinks below the optimum. The elasticities of supply and demand measure how much sellers and buyers respond to the changes in the price and, therefore, determine how much the tax distorts the market outcome. Hence, the greater the elasticities of supply and demand, the greater the deadweight loss of a tax.

SELF TEST The demand for beer is more elastic than the demand for milk. Would a tax on beer or a tax on milk have larger deadweight loss?

Deadweight Loss and Tax Revenue as Taxes Vary

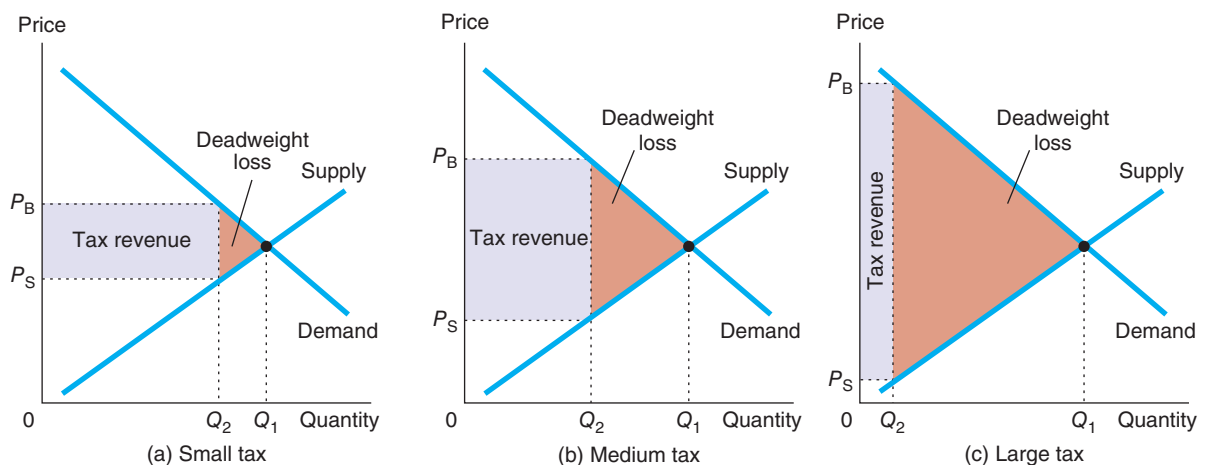
Taxes rarely stay the same for long periods of time. Policymakers are always considering raising one tax or lowering another. Here we consider what happens to the deadweight loss and tax revenue when the size of a tax changes.

Figure 9.6 shows the effects of a small, medium and large tax, holding constant the market's supply and demand curves. The deadweight loss – the reduction in total surplus that results when the tax reduces the size of a market below the optimum – equals the area of the triangle between the supply and demand curves. For the small tax in panel (a), the area of the deadweight loss triangle is quite small. But as the size of a tax rises in panels (b) and (c), the deadweight loss grows larger and larger.

FIGURE 9.6

Deadweight Loss and Tax Revenue from Three Taxes of Different Size

The deadweight loss is the reduction in total surplus due to the tax. Tax revenue is the amount of the tax times the amount of the good sold. In panel (a) a small tax has a small deadweight loss and raises a small amount of revenue. In panel (b) a somewhat larger tax has a larger deadweight loss and raises a larger amount of revenue. In panel (c) a very large tax has a very large deadweight loss, but because it has reduced the size of the market so much, the tax raises only a small amount of revenue.



Indeed, the deadweight loss of a tax rises even more rapidly than the size of the tax. The reason is that the deadweight loss is an area of a triangle, and an area of a triangle depends on the *square* of its size. If we double the size of a tax, for instance, the base and height of the triangle double, so the deadweight loss rises by a factor of 4. If we triple the size of a tax, the base and height triple, so the deadweight loss rises by a factor of 9.

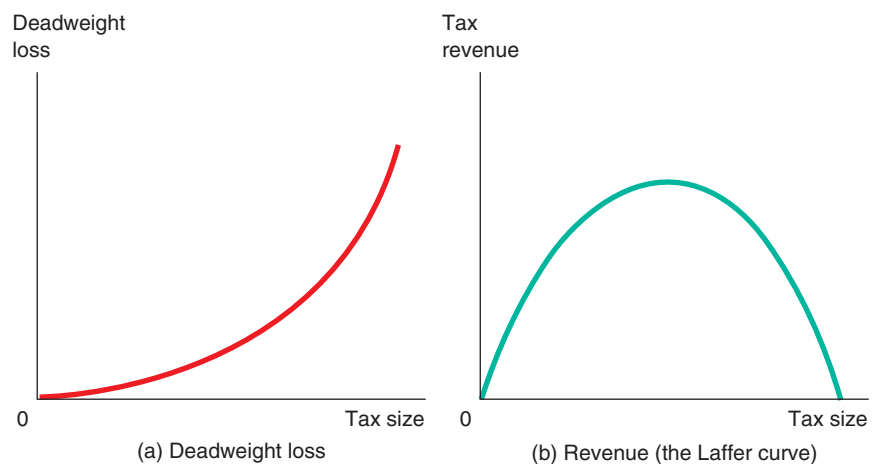
The government's tax revenue is the size of the tax times the amount of the good sold. As Figure 9.6 shows, tax revenue equals the area of the rectangle between the supply and demand curves. For the small tax in panel (a), tax revenue is small. As the size of a tax rises from panel (a) to panel (b), tax revenue grows. But as the size of the tax rises further from panel (b) to panel (c), tax revenue falls because the higher tax drastically reduces the size of the market. For a very large tax, no revenue would be raised, because people would stop buying and selling the good altogether.

Figure 9.7 summarizes these results. In panel (a) we see that as the size of a tax increases, its deadweight loss quickly gets larger. By contrast, panel (b) shows that tax revenue first rises with the size of the tax; but then, as the tax gets larger, the market shrinks so much that tax revenue starts to fall.

FIGURE 9.7

How Deadweight Loss and Tax Revenue Vary with the Size of a Tax

Panel (a) shows that, as the size of a tax grows larger, the deadweight loss grows larger. Panel (b) shows that tax revenue first rises, then falls. This relationship is sometimes called the Laffer curve.



SELF TEST If the government doubles the tax on petrol, can you be sure that revenue from the petrol tax will rise? Can you be sure that the deadweight loss from the petrol tax will rise? Explain.

ADMINISTRATIVE BURDEN

The second cost that a tax imposes is administrative burden and a well-designed tax system seeks to reduce this to a minimum. In many countries, individuals and businesses are required to alert the tax authorities of their incomes and business activities so that the correct tax can be collected. These forms can often be long and complex and for businesses, in particular, this can be a time consuming and stressful operation. If you ask the typical person required to do this for an opinion about the tax system, the opinion is unlikely to be favourable. The administrative burden of any tax system is part of the inefficiency it creates. This burden includes not only the time spent filling out forms but also the time spent throughout the year keeping records for tax purposes and the resources the government has to use to enforce the tax laws.

Many taxpayers hire accountants or tax lawyers to help them with their taxes. These experts in the complex tax laws fill out the tax forms for their clients and help clients arrange their affairs in a way that

reduces the amount of taxes owed. This behaviour is avoidance (optimizing your affairs so that you pay as little tax as possible without breaking the law, which is different from tax evasion – i.e. lying about your affairs in order to reduce the amount of tax paid). Tax avoidance is perfectly legal. Tax evasion is illegal.

Critics of the tax system say that these advisors help their clients avoid taxes by abusing some of the detailed provisions of the tax system (or, as it is sometimes termed, the tax code). These detailed provisions are often dubbed 'loopholes'. In some cases, loopholes are government mistakes: they arise from ambiguities or omissions in the tax laws. More often, they arise because the government has chosen to give special treatment to specific types of behaviour. For example, the UK tax system allows money spent on a personal pension plan to be exempt from income tax, up to a certain limit. This is because the government wants to encourage people to save for their retirement.

The resources devoted to complying with the tax laws are a type of deadweight loss. The government gets only the amount of taxes paid. By contrast, the taxpayer loses not only this amount but also the time and money spent documenting, computing and avoiding taxes.

The administrative burden of the tax system could be reduced by simplifying the tax laws. Yet simplification is often politically difficult. Most people are ready to simplify the tax code by eliminating the loopholes that benefit others, yet few are eager to give up the loopholes that they use. In the end, the complexity of tax law results from the political process as various taxpayers with their own special interests lobby for their causes. This process is called 'rent seeking' and is part of a branch of economics called public choice theory. Public choice theory is about the analysis of governmental behaviour, and the behaviour of individuals who interact with government.

THE DESIGN OF THE TAX SYSTEM

There are a number of key factors in the design of a good tax system. Many economists would agree that there are some fundamental principles that characterize a good tax system but the reality is that most countries have very complex tax systems which might compromise both the efficiency of the system and the desire to be equitable. We are going to look at some of these principles in the next section.

Adam Smith's Four Canons of Taxation

The 18th century economist Adam Smith suggested that any good tax system should adhere to four basic principles or canons. These four principles are:

- **Equality** – that each person should pay taxes according to their ability to pay so that the rich should pay more in taxes than the poor
- **Certainty** – tax payers need to know what taxes they are liable for and be able to plan ahead on this basis and at the same time governments should be able to have some certainty in how much they are able to collect in taxes
- **Convenience** – paying taxes should be made as easy as possible and tax systems should be designed to be as simple as possible in order to help maximize tax revenue
- **Economic** – any tax system must ensure that the cost of collecting and administering taxes is less than the amount collected.

Whilst these principles provide some guidance in the design of a good tax system they raise many questions. For example, if we agree on the principle that the rich should pay more in taxes than the poor, how much more should they pay? How do we define 'rich'? is it fair that someone who has worked hard should pay far more in taxes than someone who is inherently lazy? At what point will this principle lead people to begin to seek ways to avoid tax and possibly exit the tax system altogether because of high tax rates?

We will look in more detail at some of the issues surrounding the design of a good tax system later in this section but we are now going to introduce some key concepts in any tax system.

Marginal Tax Rates versus Average Tax Rates

When discussing the efficiency and equity of income or direct taxes, economists distinguish between two notions of the tax rate: the average and the marginal. The **average tax rate** (ATR) is total taxes paid divided by total income and can be expressed by the formula:

$$\text{ATR} = \frac{\text{Tax liability}}{\text{Taxable income}}$$

Where tax liability is the amount that the individual is liable to pay in tax to the tax authorities – the total tax paid.

average tax rate total taxes paid divided by total income

The **marginal tax rate** (MTR) is the extra taxes paid on an additional unit of income expressed by the formula:

$$\text{MTR} = \frac{\text{Change in tax liability}}{\text{Change in taxable income}}$$

For example, suppose that the government taxes 20 per cent of the first €50,000 of income and 50 per cent of all income above €50,000. Under this tax, a person who earns €60,000 pays tax of €15,000: 20 per cent of the first €50,000 ($0.20 \times €50,000 = €10,000$) plus 50 per cent of the next €10,000 ($0.50 \times €10,000 = €5,000$). For this person, the average tax rate is €15,000/€60,000, or 25 per cent. But the marginal tax rate is 50 per cent. If the taxpayer earned an additional euro of income, that euro would be subject to the 50 per cent tax rate, so the amount the taxpayer would owe to the government would rise by €0.50.

marginal tax rate the extra taxes paid on an additional unit of income

The marginal and average tax rates each contain a useful piece of information. If we are trying to gauge the sacrifice made by a taxpayer, the average tax rate is more appropriate because it measures the fraction of income paid in taxes. By contrast, if we are trying to gauge how much the tax system distorts incentives, the marginal tax rate is more meaningful. One of the *Ten Principles of Economics* is that rational people think at the margin. A corollary to this principle is that the marginal tax rate measures how much the tax system discourages people from working. If you are thinking of working an extra few hours, the marginal tax rate determines how much the government takes of your additional earnings. It is the marginal tax rate, therefore, that determines the deadweight loss of an income tax.

Lump-Sum Taxes

Suppose the government imposes a tax of €4,000 on everyone. That is, everyone owes the same amount, regardless of earnings or any actions that a person might take. Such a tax is called a **lump-sum tax**.

lump-sum tax a tax that is the same amount for every person

A lump-sum tax shows clearly the difference between average and marginal tax rates. For a taxpayer with income of €20,000, the average tax rate of a €4,000 lump-sum tax is 20 per cent; for a taxpayer with income of €40,000, the average tax rate is 10 per cent. For both taxpayers, the marginal tax rate is zero because no tax is owed on an additional unit of income.

A lump-sum tax is the most efficient tax possible. Because a person's decisions do not alter the amount owed, the tax does not distort incentives and, therefore, does not cause deadweight losses. Because everyone can easily compute the amount owed and because there is no benefit to hiring tax lawyers and accountants, the lump-sum tax imposes a minimal administrative burden on taxpayers. If lump-sum taxes are so efficient, why do we rarely observe them in the real world? The reason is that efficiency is only one goal of the tax system. A lump sum tax would take the same amount from the poor and the rich, an outcome most people would view as unfair and contrary to Smith's first canon. In the next section we consider the issue of equity.

SELF TEST What is meant by the efficiency of a tax system? • What can make a tax system inefficient?

CASE STUDY The Deadweight Loss Debate

The ideas covered so far in this chapter go to the heart of a profound political question: how big should the government sector be? The debate hinges on these concepts because the larger the deadweight loss of taxation, the larger the welfare cost of any government expenditure programme such as providing public health services or national defence. If taxation entails large deadweight losses, then these losses are a strong argument for a leaner government that does less and taxes less. But if taxes impose small deadweight losses, then government programmes are less costly than they otherwise might be.

So how big are the deadweight losses of taxation? This is a question about which economists disagree. To see the nature of this disagreement, consider one of the most important taxes in most advanced economies – the tax on labour. In the UK, for example, National Insurance contributions and, to a large extent, income tax, are taxes on labour. A labour tax places a wedge between the wage that firms pay and the wage that workers receive. If we add both forms of labour taxes together, the *marginal tax rate* on labour income – the tax on the last pound of earnings – is around 33 per cent for most UK manual workers. In some European countries – particularly Scandinavian countries – the marginal rate is even higher.

Although the size of the labour tax is easy to determine, the deadweight loss of this tax is less straightforward. Economists disagree about whether this 33 per cent labour tax has a small or a large deadweight loss. This disagreement arises because economists hold different views about the elasticity of labour supply.

Economists who argue that labour taxes are not very distorting believe that labour supply is fairly inelastic. Most people, they claim, would work full-time regardless of the wage. If so, the labour supply curve is almost vertical, and a tax on labour has a small deadweight loss.

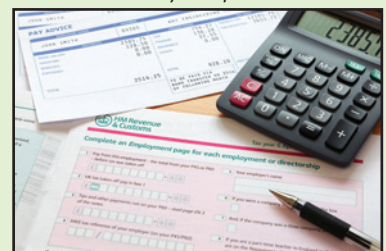
Economists who argue that labour taxes are highly distorting believe that labour supply is more elastic. They admit that some groups of workers may supply their labour inelastically but claim that many other groups respond more to incentives. Here are some examples:

- Many workers can adjust the number of hours they work – for instance, by working overtime. The higher the wage, the more hours they choose to work.
- Some families have second earners – often married women with children – with some discretion over whether to do unpaid work at home or paid work in the marketplace. When deciding whether to take a job, these second earners compare the benefits of being at home (including savings on the cost of child care) with the wages they could earn.
- Many people can choose when to retire, and their decisions are partly based on the wage. Once they are retired, the wage determines their incentive to work part-time.
- Some people consider engaging in illegal economic activity, such as the drug trade, or working at jobs that pay ‘under the table’ to evade taxes. Economists call this the *black economy* (or sometimes the underground economy). In deciding whether to work in the black economy or at a legitimate job, these potential criminals compare what they can earn by breaking the law with the wage they can earn legally.

In each of these cases, the quantity of labour supplied responds to the wage (the price of labour). Thus, the decisions of these workers are distorted when their labour earnings are taxed. Labour taxes encourage workers to work fewer hours, second earners to stay at home, the elderly to retire early and the unscrupulous to enter the black economy.

These two views of labour taxation persist to this day. Indeed, whenever you see two political candidates debating whether the government should provide more services or reduce the tax burden, keep in mind that part of the disagreement may rest on different views about the elasticity of labour supply and the deadweight loss of taxation.

The size of the deadweight loss of the tax system is much debated but few deny that the tax systems in most countries are very complex.



TAXES AND EQUITY

In any country, tax policy always generates some of the most heated political debates. The heat is rarely fuelled by questions of efficiency. Instead, it usually arises from disagreements over how the tax burden should be distributed. Of course, if we are to rely on the government to provide some of the goods and services we want, taxes must fall on someone. How should the burden of taxes be divided among the population? How do we evaluate whether a tax system is fair? Everyone agrees that the tax system should be equitable, but there is much disagreement about what equity means and how the equity of a tax system can be judged.

The Benefits Principle

One principle of taxation, called the **benefits principle**, states that people should pay taxes based on the benefits they receive from government services. It seems fair that a person who often goes to the cinema pays more in total for cinema tickets than a person who rarely goes. Similarly, a person who gets great benefit from a good provided by the state should pay more for it than a person who gets little benefit.

benefits principle the idea that people should pay taxes based on the benefits they receive from government services

The duty on petrol, for instance, is sometimes justified using the benefits principle, since in some countries, revenues from the tax on petrol are used to build and maintain roads. Because those who buy petrol are the same people who use the roads, this tax might be viewed as a fair way to pay for this government service.

The benefits principle can also be used to argue for Smith's first canon – that wealthy citizens should pay higher taxes than poorer ones. Why? Simply because the wealthy benefit more from public services. Consider, for example, the benefits of police protection from theft. Citizens with much to protect, it could be argued, get greater benefit from the police than do those with less to protect. Therefore, according to the benefits principle, the wealthy should contribute more than the poor to the cost of maintaining the police force. The same argument can be used for many other public services, such as fire protection, national defence and the criminal justice system.

It is even possible to use the benefits principle to argue for anti-poverty programmes funded by taxes on the wealthy. If we assume that people prefer living in a society without poverty, this suggests that anti-poverty programmes are a desirable policy objective. If the wealthy place a greater value on this objective than other members of society, perhaps just because the wealthy have more to spend, then, according to the benefits principle, they should be taxed more heavily to pay for these programmes.

The Ability-to-Pay Principle

Another way to evaluate the equity of a tax system is called the **ability-to-pay principle**, which states that taxes should be levied on a person according to how well that person can shoulder the burden. This principle is sometimes justified by the claim that all citizens should make an 'equal sacrifice' to support the government. The magnitude of a person's sacrifice, however, depends not only on the size of his tax payment but also on his income and other circumstances. A €1,000 tax paid by a poor person may require a larger sacrifice than a €10,000 tax paid by a rich one.

ability-to-pay principle the idea that taxes should be levied on a person according to how well that person can shoulder the burden

The ability-to-pay principle leads to two corollary notions of equity: vertical equity and horizontal equity.

Vertical equity states that taxpayers with a greater ability to pay taxes should contribute a larger amount.

Horizontal equity states that taxpayers with similar abilities to pay should contribute the same amount.

vertical equity the idea that taxpayers with a greater ability to pay taxes should pay larger amounts

horizontal equity the idea that taxpayers with similar abilities to pay taxes should pay the same amount

Although these notions of equity are widely accepted, applying them to evaluate a tax system is rarely straightforward.

Vertical Equity If taxes are based on ability to pay, then richer taxpayers should pay more than poorer taxpayers. But how much more should the rich pay? Much of the debate over tax policy concerns this question.

Consider the three tax systems in Table 9.1. In each case, taxpayers with higher incomes pay more. Yet the systems differ in how quickly taxes rise with income. The first system is called a **proportional tax** or sometimes a **flat tax** because all taxpayers pay the same fraction of income. The second system is called **regressive** because high-income taxpayers pay a smaller fraction of their income, even though they pay a larger amount. The third system is called **progressive** because high-income taxpayers pay a larger fraction of their income.

proportional or flat tax a tax for which high-income and low-income taxpayers pay the same fraction of income
regressive tax a tax for which high-income taxpayers pay a smaller fraction of their income than do low-income taxpayers
progressive tax a tax for which high-income taxpayers pay a larger fraction of their income than do low-income taxpayers

Which of these three tax systems is most fair? Sometimes it is argued that a progressive tax system is fairer because richer people pay more tax and they can afford to do so because they are richer. But richer people will also pay more tax than poorer people under a flat tax system or even under a regressive tax system. In fact, there is no obvious answer, and economic theory does not offer any help in trying to find one. Equity, like beauty, is in the eye of the beholder.

TABLE 9.1**Three Tax Systems**

| Income | Proportional or flat tax | | Regressive tax | | Progressive tax | |
|---------|--------------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|
| | Amount of tax (€) | Percentage of income (%) | Amount of tax (€) | Percentage of income (%) | Amount of tax (€) | Percentage of income (%) |
| €50,000 | 12,500 | 25 | 15,000 | 30 | 10,000 | 20 |
| 100,000 | 25,000 | 25 | 25,000 | 25 | 25,000 | 25 |
| 200,000 | 50,000 | 25 | 40,000 | 20 | 60,000 | 30 |

Horizontal Equity If taxes are based on ability to pay, then similar taxpayers should pay similar amounts of taxes. But what determines if two taxpayers are similar? Individuals' circumstances can differ in many ways. To evaluate whether a tax system is horizontally equitable, one must determine which differences are relevant for a person's ability to pay and which differences are not.

Suppose Mr Smith and Ms Jones each have an income of €50,000 a year. Mr Smith is unmarried and has no children, but he has an illness that means that he has to employ someone to be with him while he is at work and getting to and from his place of business. This costs him €20,000 a year. Ms Jones is in good health and is a lone parent with a child aged three. Ms Jones has to pay €15,000 a year for child care while she is at work. Would it be fair for Mr Smith and Ms Jones to pay the same tax because they have the same income? Would it be fairer to give Mr Smith a tax break to help him offset the costs of a caring assistant? Would it be fairer to give Ms Jones a tax break to compensate her for the cost of childcare?

There are no easy answers to these questions.

Tax Incidence and Tax Equity

Tax incidence – the study of who bears the burden of taxes – is central to evaluating tax equity. We know from Chapter 8 that the person who bears the burden of a tax is not always the person who gets the tax bill from the government. Because taxes alter supply and demand, they alter equilibrium prices. As a

result, they affect people beyond those who, according to statute, actually pay the tax. When evaluating the vertical and horizontal equity of any tax, it is important to take account of these indirect effects.

Many discussions of tax equity ignore the indirect effects of taxes and are based on what economists mockingly call the flypaper theory of tax incidence. According to this theory, the burden of a tax, like a fly on flypaper, sticks wherever it first lands. This assumption, however, is rarely valid.

For example, a person not trained in economics might argue that a tax on expensive fur coats is vertically equitable because most buyers of furs are wealthy. Yet if these buyers can easily substitute other luxuries for furs, then a tax on furs might only reduce the sale of furs. In the end, the burden of the tax will fall more on those who make and sell furs than on those who buy them. Because most workers who make furs are not wealthy, the equity of a fur tax could be quite different from what the flypaper theory indicates.

Corporate taxes – taxes levied on the profits that businesses make – provides a good example of the importance of tax incidence for tax policy. Corporate taxes are popular among voters. After all, corporations are not people. Voters are always eager to have their taxes reduced and have some impersonal corporation pick up the tab. In a sense, corporation tax appears to be a tax that is imposed on nobody. No wonder it's popular with policymakers also.

But before deciding that corporation tax is a good way for the government to raise revenue, we should consider who bears the burden of corporation tax. This is a difficult question on which economists disagree, but one thing is certain: *people pay all taxes*. When the government levies a tax on a corporation, the corporation is more like a tax collector than a taxpayer. The burden of the tax ultimately falls on people – the owners, customers or workers of the corporation.

Many economists believe that workers and customers bear much of the burden of corporation tax. To see why, consider an example. Suppose that the government decides to raise the tax on the income earned by car companies. At first, this tax hurts the owners of the car companies who receive less profit. But, over time, these owners will respond to the tax. Because producing cars is less profitable, they invest less in building new car factories. Instead, they invest their wealth in other ways – for example, by buying larger houses or by building factories in other industries or other countries. With fewer car factories, the supply of cars declines, as does the demand for car workers. Thus, a tax on corporations making cars causes the price of cars to rise and the wages of car workers to fall.

Corporation tax shows how dangerous the flypaper theory of tax incidence can be. Corporation tax is popular in part because it appears to be paid by rich corporations. Yet those who bear the ultimate burden of the tax – the customers and workers of corporations – are often not rich. If the true incidence of corporation tax were more widely known, this tax might be less popular among voters – and policymakers.

SELF TEST Explain the benefits principle and the ability-to-pay principle. ● What are vertical equity and horizontal equity? ● Why is studying tax incidence important for determining the equity of a tax system?

CONCLUSION

Taxes, someone once said, are the price we pay for a civilized society. Indeed, our society cannot exist without some form of taxes. We all expect the government to provide us with certain services, such as roads, parks, police and national defence. These public services require tax revenue. This chapter has shed some light on how high the price of civilized society can be. One of the *Ten Principles of Economics* is that markets are usually a good way to organize economic activity. When the government imposes taxes on buyers or sellers of a good, however, society loses some of the benefits of market efficiency. Taxes are costly to market participants not only because taxes transfer resources from those participants to the government, but also because they alter incentives and distort market outcomes. Throughout this chapter we have assumed that the benefit to the government, or more accurately, the third party who receives the benefit of the tax spending, is equal to the tax revenue given by $T \times Q$. This may not always be the case. To get a more accurate analysis of the size of deadweight losses arising from taxes, we may have to find

a way to place a value on the benefits of the tax spending. For example, if taxes are spent on improving the road infrastructure there is a benefit to car drivers which may be greater than the tax spending. Calculating the value of these benefits is not easy but has to be carried out to get a more accurate picture of the change in welfare. How the government chooses to spend tax revenue is also a factor to consider. A paper written by Gupta, Verhoeven and Tiongson (Gupta, S. Verhoeven, M. and Tiongson, E.R. (2004) 'Public Spending on Health Care and the Poor', in Gupta, S. Clements, B., Inchauste, G. (eds), *Helping Countries Develop: The Role of Fiscal Policy*, Washington, DC: IMF) concluded that in 70 developing and transition economies there was evidence to suggest that the impact of public spending on health care on the poor could be substantial in comparison with spending on the non-poor. A 1 per cent increase in public spending on health, for example, reduces child mortality by twice as many deaths among the poor as the non-poor and infant mortality rates are similarly affected. This highlights that there are many other factors that may have to be taken into consideration when analysing changes in welfare as a result of taxes.

In the second half of this chapter we looked at equity and efficiency in tax systems. Almost everyone agrees that equity and efficiency are the two most important goals of the tax system. But often these two goals conflict. Many proposed changes in the tax laws increase efficiency while reducing equity, or increase equity whilst reducing efficiency. People disagree about tax policy often because they attach different weights to these two goals.

Economics alone cannot determine the best way to balance the goals of efficiency and equity. This issue involves political philosophy as well as economics. But economists do have an important role in the political debate over tax policy: they can shed light on the trade-offs that society faces and can help us avoid policies that sacrifice efficiency without any benefit in terms of equity.

IN THE NEWS

Tax Avoidance

As many countries in Europe struggle to come to terms with managing public finances, the efficiency and equity of tax systems has come under some scrutiny. One particular issue that aroused considerable anger and debate is corporate tax avoidance.

Avoiding Tax – Good Business or Moral Bankruptcy?

In 2012 reports surfaced in the UK that some high profile corporations operating in the country were not paying any corporate taxes. Three companies in particular were singled out for particular ire: Google, Starbucks and Amazon. The way that these companies had been managing their accounts meant that they were able to reduce their liabilities for any corporation tax in the UK. On the face of it this seemed bizarre,

after all, Starbucks was reported to have generated almost £400 million in revenue in the UK in 2011 but paid no corporation tax and over a 13-year period generated £3.1 billion in sales paying £8.6 million in tax. Google was reported to have generated £2.5 billion in sales in the UK in 2011, with an overall group profit margin of 33 per cent, but its UK operation paid £3.4 million in 2011. Amazon paid £1.8 million in tax on a turnover of £207 million in the same year. According to the way these firms' accounts are prepared, they are not generating the profits on

which tax is payable. In Starbucks, case, the company claimed it had made a loss in the 15 years it has been operating in the UK apart from 2006.

In each case it is not that the companies are making overall losses and therefore are not liable for corporation tax but the way in which they record the various cost and revenue flows they receive. Google, for example, records revenues it receives in Dublin but then has to pay royalty payments to a subsidiary it owns in the Netherlands and the remainder is then reported as profit

(Continued)

in another subsidiary in Bermuda where it does not attract tax. Amazon said that its UK operations were actually just back-office functions for its main business in Europe based in Luxembourg and Starbucks noted that whilst it had made profits on sales in the UK, when these profits were set against exceptional capital items it actually made a loss.

News reports had to make it very clear that none of these companies was doing anything illegal and that tax avoidance was simply a case of the companies using the rules of the tax system to their advantage. Indeed, Google's chairman, Eric Schmidt, was quoted as saying "It's called capitalism ... We are proudly capitalistic. I'm not confused about this."

"We pay lots of taxes; we pay them in the legally prescribed ways. I am very proud of the structure that we set up. We did it based on the incentives that the governments offered us to operate."

The anger at the behaviour of these firms was widespread. Radio phone-ins were not short of people complaining of the moral vacuum in which corporations operated and that these organizations ought to pay their fair share of tax like everyone else has to. Demonstrations outside some Starbucks outlets and the strength of the feedback the

company received from customers prompted it to announce that it was going to volunteer to pay £20 million to the UK tax authorities between 2013 and 2014 – taxes, it was quick to point out, it did not legally have to pay.

Representatives of the companies and some commentators who sought to defend them noted that in fact these three firms pay lots of taxes and contribute to the UK economy in so many ways. They employ thousands of workers, for example, all of whom pay labour taxes. They contribute millions of pounds in value added tax and in investing in the UK economy provide wider economic benefits.

Some of the supporters of the firms noted that it was up to the government to get its house in order and to close tax loopholes which allow firms to set up often complex schemes to avoid tax. Indeed, across Europe there were calls for governments to focus on doing just that to help reduce tax avoidance across the European Union' which is costing European governments billions in lost tax revenue. The issue was scheduled to be on the agenda at meetings of the developed countries' finance minister meetings throughout 2013. One suggestion which has been made and which has been supported by a report coming out of the

Centre for Business Taxation at the University of Oxford in June 2012, is to reduce the marginal rate of taxation for companies, which would in turn lead to more of an incentive for businesses to pay tax.

Questions

- 1 Distinguish between tax avoidance and tax evasion.
- 2 Many of the schemes used to avoid tax are complex and require the employment of tax experts from accountancy firms. What conclusions can you arrive at about the relative costs and benefits of firms engaging in these complex schemes?
- 3 What do you think the Google chairman Eric Schmidt meant when he said 'It's called capitalism ...'?
- 4 Given that the three companies highlighted in this article contribute a considerable amount to the UK economy in other ways, to what extent would you agree with the view that a few million in avoided corporation tax is a price worth paying to have them in the UK economy?
- 5 How does this article highlight the need for a well-designed tax system and also the need to trade-off efficiency and equity in any tax system?



Amazon, Google and Starbucks attracted a good deal of publicity about the amount of tax they pay in the UK. The issue, however, is much more complex than one of tax avoidance. Companies such as these might bring many benefits to the UK economy.

SUMMARY

- The efficiency of a tax system refers to the costs that it imposes on taxpayers. There are two costs of taxes beyond the transfer of resources from the taxpayer to the government. The first is the distortion in the allocation of resources that arises as taxes alter incentives and behaviour. The second is the administrative burden of complying with the tax laws.
- A tax on a good reduces the welfare of buyers and sellers of the good, and the reduction in consumer and producer surplus usually exceeds the revenue raised by the government. The fall in total surplus – the sum of consumer surplus, producer surplus and tax revenue – is called the deadweight loss of the tax.
- Taxes have deadweight losses because they cause buyers to consume less and sellers to produce less, and this change in behaviour shrinks the size of the market below the level that maximizes total surplus. Because the elasticities of supply and demand measure how much market participants respond to market conditions, larger elasticities imply larger deadweight losses.
- As a tax grows larger, it distorts incentives more, and its deadweight loss grows larger. Tax revenue first rises with the size of a tax. Eventually, however, a larger tax reduces tax revenue because it reduces the size of the market.
- The equity of a tax system concerns whether the tax burden is distributed fairly among the population. According to the benefits principle, it is fair for people to pay taxes based on the benefits they receive from the government. According to the ability-to-pay principle, it is fair for people to pay taxes based on their capability to handle the financial burden. When evaluating the equity of a tax system, it is important to remember a lesson from the study of tax incidence: the distribution of tax burdens is not the same as the distribution of tax bills.
- When considering changes in the tax laws, policy-makers often face a trade-off between efficiency and equity. Much of the debate over tax policy arises because people give different weights to these two goals.

QUESTIONS FOR REVIEW

- 1 What are the two main reasons that governments levy taxes?
- 2 What happens to consumer and producer surplus when the sale of a good is taxed? How does the change in consumer and producer surplus compare to the tax revenue? Explain.
- 3 Draw a supply and demand diagram with a tax on the sale of the good. Show the deadweight loss. Show the tax revenue.
- 4 How do the elasticities of supply and demand affect the deadweight loss of a tax? Why do they have this effect?
- 5 What happens to the deadweight loss and tax revenue when a tax is increased?
- 6 What are the four canons of taxation?
- 7 Explain how corporate profits are taxed twice.
- 8 Why is the burden of a tax to taxpayers greater than the revenue received by the government?
- 9 Give two arguments why wealthy taxpayers should pay more taxes than poor taxpayers.
- 10 What is the concept of horizontal equity, and why is it hard to apply?

PROBLEMS AND APPLICATIONS

- 1 The market for pizza is characterized by a downwards sloping demand curve and an upwards sloping supply curve.
 - a. Draw the competitive market equilibrium. Label the price, quantity, consumer surplus and producer surplus. Is there any deadweight loss? Explain.
 - b. Suppose that the government forces each pizzeria to pay a €11 tax on each pizza sold. Illustrate the effect of this tax on the pizza market, being sure to label the consumer surplus, producer surplus, government revenue and deadweight loss. How does each area compare to the pre-tax case?
 - c. If the tax was removed, pizza eaters and sellers would be better off, but the government would lose tax revenue. Suppose that consumers and producers voluntarily transferred some of their gains to the government. Could all parties (including the government) be better off than they were with a tax?
 - d. Explain using the labelled areas in your graph.
- 2 Evaluate the following two statements. Do you agree? Why or why not?
 - a. 'If the government taxes land, wealthy landowners will pass the tax on to their poorer renters.'
 - b. 'If the government taxes blocks of flats, wealthy landlords will pass the tax on to their poorer renters.'

3 Evaluate the following two statements. Do you agree? Why or why not?

- 'A tax that has no deadweight loss cannot raise any revenue for the government.'
- 'A tax that raises no revenue for the government cannot have any deadweight loss.'

4 Consider the market for rubber bands.

- If this market has very elastic supply and very inelastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers? Use the tools of consumer surplus and producer surplus in your answer.
- If this market has very inelastic supply and very elastic demand, how would the burden of a tax on rubber bands be shared between consumers and producers?
- Contrast your answer with your answer to part (a).

5 Suppose that the government imposes a tax on heating oil.

- Would the deadweight loss from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.
- Would the revenue collected from this tax likely be greater in the first year after it is imposed or in the fifth year? Explain.

6 This chapter analysed the welfare effects of a tax on a good. Consider now the opposite policy. Suppose that the government subsidizes a good: for each unit of the good sold, the government pays €2 to the buyer. How does the subsidy affect consumer surplus, producer surplus, tax revenue and total surplus? Does a subsidy lead to a deadweight loss? Explain.

7 A sales tax of 20 per cent is levied on clothing for adults but not on children's clothing. Discuss the merits of this distinction, considering both efficiency and equity.

8 Suppose that the tax system had the following features. Explain how individuals' behaviour is affected.

- Contributions to charity are tax-deductible.
- Sales of beer are taxed.

c. Realized capital gains are taxed, but accrued gains are not. (When someone owns a share of stock that rises in value, she has an 'accrued' capital gain. If she sells the share, she has a 'realized' gain.)

9 Suppose that a market is described by the following supply and demand equations:

$$Q_s = 2P$$

$$Q_d = 300 - P$$

- Solve for the equilibrium price and the equilibrium quantity.
- Suppose that a tax of T is placed on buyers, so the new demand equation is:

$$Q_d = 300 - (P + T):$$

Solve for the new equilibrium. What happens to the price received by sellers, the price paid by buyers and the quantity sold?

- Tax revenue is $T \times Q$. Use your answer to part (b) to solve for tax revenue as a function of T . Graph this relationship for T between 0 and 300.
- The deadweight loss of a tax is the area of the triangle between the supply and demand curves. Recalling that the area of a triangle is $1/2 \times \text{base} \times \text{height}$, solve for deadweight loss as a function of T . Graph this relationship for T between 0 and 300. (Hint: looking sideways, the base of the deadweight loss triangle is T , and the height is the difference between the quantity sold with the tax and the quantity sold without the tax.)
- The government now levies a tax on this good of €200 per unit. Is this a good policy? Why or why not? Can you propose a better policy?

10 The tax on cigarettes and other smoking products is very high in many countries (in the UK it is over 80 per cent of the selling price) and has been rising over time. Discuss the merits of this policy, considering the principles of equity and efficiency.

PART 5

INEFFICIENT MARKET ALLOCATIONS

10 PUBLIC GOODS, COMMON RESOURCES AND MERIT GOODS

In all the chapters so far we have talked about the market providing goods and services. Demand reflects the benefit and value buyers put on goods and supply reflects the cost to producers. Price acts as a signal to buyers and sellers and allocates scarce resources among competing uses. Based on this understanding, how much would you, individually, be prepared to pay for national defence, justice or a police service? Would you be prepared to pay for street lighting, local playgrounds and parks? Have you ever thought about who owns nature – things such as rivers, mountains, beaches, lakes and oceans? Many people enjoy the benefit of these goods but do not pay directly for them. These goods and services come under special categories in economic analysis. When goods are available but users do not have to pay directly, market forces that normally allocate resources in our economy are absent.

These types of goods would either not be produced under a market system or not produced in quantities deemed appropriate to satisfy needs. When this happens, the market fails to allocate scarce resources efficiently (because needs are not being met) and as a result governments step in to remedy this so-called market failure. This is one of the *Ten Principles of Economics*: governments can sometimes improve market outcomes. This chapter and the next will look at market failure in more detail. When a good does not have a price attached to it, private markets cannot ensure that the good is produced and consumed in the proper amounts. In such cases, government policy can potentially remedy the market failure and raise economic well-being.

When governments intervene to provide goods and services, these goods and services are provided by the public sector. The **public sector** refers to the provision of goods and services which are funded and organized by the state on behalf of the population as a whole. This contrasts with the **private sector** where goods and services are funded and organized by private firms, usually, but not exclusively, with a view to making a profit.

public sector that part of the economy where business activity is owned, financed and controlled by the state, and goods and services are provided by the state on behalf of the population as a whole

private sector that part of the economy where business activity is owned, financed and controlled by private individuals

THE DIFFERENT KINDS OF GOODS

How well do markets work in providing the goods that people want? The answer to this question depends on the good being considered. In many cases, markets can be relied upon to provide efficient outcomes: the price of goods adjusts to balance supply and demand, and this equilibrium maximizes the sum of producer and consumer surplus. With some goods, however, this outcome cannot be assumed. In thinking about the various goods in the economy, it is useful to group them according to two characteristics:

- Is the good **excludable**? Can people who do not pay for the use of a good be prevented from using the good?
- Is the good **rival**? Does one person's use of the good diminish another person's ability to use it?

excludable the property of a good whereby a person can be prevented from using it when they do not pay for it
rival the property of a good whereby one person's use diminishes other people's use

Using these two characteristics, Figure 10.1 divides goods into four categories:

- 1. Private goods** are both excludable and rival. Consider a chocolate bar, for example. A chocolate bar is excludable because it is possible to prevent someone from eating the bar – you just don't give it to him. A chocolate bar is rival because if one person eats the bar, another person cannot eat the same bar. Most goods in the economy are private goods like chocolate bars. Our analysis of supply and demand so far has been based on the implicit assumption that goods were both excludable and rival.
- 2. Public goods** are neither excludable nor rival. That is, people cannot be prevented from using a public good, and one person's use of a public good does not reduce another person's ability to use it. For example, a country's national defence system: it protects all of the country's citizens equally and the fact that one person is being defended does not affect whether or not another citizen is defended.

private goods goods that are both excludable and rival
public goods goods that are neither excludable nor rival

- 3. Common resources** are rival but not excludable. For example, fish in the ocean are a rival good: when one person catches fish, there are fewer fish for the next person to catch. Yet these fish are not an excludable good because, given the vast size of an ocean, it is difficult to stop fishermen from taking fish out of it when, for example, they have not paid for a licence to do so.

common resources goods that are rival but not excludable

- 4. Natural monopoly** When a good is excludable but not rival, it is an example of a natural monopoly. For instance, consider fire protection in a small town. It is easy to exclude people from using this good: the fire service can just let their houses burn down. Yet fire protection is not rival. Fire fighters spend much of their time waiting for a fire, so protecting an extra house is unlikely to reduce the protection available to others. In other words, once a town has paid for the fire service, the additional cost of protecting one more house is small. We will look at natural monopolies in more detail in a later chapter.

natural monopoly a good which is excludable but non-rival

FIGURE 10.1**Four Types of Goods**

Goods can be grouped into four categories according to two questions: (1) Is the good excludable? That is, can people be prevented from using it? (2) Is the good rival? That is, does one person's use of the good diminish other people's use of it? This diagram gives examples of goods in each of the four categories.

| | | Rival? | |
|-------------|-----|--|--|
| | | Yes | No |
| Excludable? | Yes | Private goods <ul style="list-style-type: none"> • Chocolate bars • Clothing • Congested toll roads | Natural monopolies <ul style="list-style-type: none"> • The fire service • Cable TV • Uncongested toll roads |
| | No | Common resources <ul style="list-style-type: none"> • Fish in the ocean • The environment • Congested non-toll roads | Public goods <ul style="list-style-type: none"> • Flood-control dams • National defence • Uncongested non-toll roads |

For both public goods and common resources, a cost or benefit arises because something of value has no price attached to it. If one person were to provide a public good, such as a national defence system, other people would be better off, and yet they could not be charged for this benefit. Similarly, when one person uses a common resource, such as the fish in the ocean, other people are worse off, and yet they are not compensated for this loss. These are referred to as *external effects* or *externalities* and we will look at externalities in more detail in Chapter 11. Because of these external effects, private decisions about consumption and production can lead to an inefficient allocation of resources, and government intervention can potentially raise economic well-being.

SELF TEST Define *public goods* and *common resources*, and give an example of each.

PUBLIC GOODS

To understand how public goods differ from other goods and what problems they present for society, let's consider an example: a fireworks display. This good is not excludable because it is impossible to prevent someone from seeing fireworks, and it is not rival because one person's enjoyment of fireworks does not reduce anyone else's enjoyment of them.

The Free Rider Problem

The citizens of a small Spanish town, Hereza, like seeing fireworks on 6 January when Spain celebrates Epiphany. Each of the town's 500 residents places a €10 value on the experience. The cost of putting on a fireworks display is €1,000. Because the €5,000 of benefits exceed the €1,000 of costs, it is efficient for Hereza residents to have a fireworks display on 6 January.

Would the private market produce the efficient outcome? Probably not. Imagine that Conchita, a Hereza entrepreneur, decided to put on a fireworks display. Conchita would surely have trouble selling tickets to the event because her potential customers would quickly figure out that they could see the fireworks even without a ticket (unless Conchita could figure out a way in which she could keep the display 'private' somehow). Because fireworks are not excludable, people have an incentive to be free riders. A **free rider** is a person who receives the benefit of a good but avoids paying for it.

free rider a person who receives the benefit of a good but avoids paying for it

One way to view this market failure is that it arises because of an externality. If Conchita did put on the fireworks display, she would confer an external benefit on those who saw the display without paying for it. When deciding whether to put on the display, Conchita ignores these external benefits. Even though a fireworks display is socially desirable, it is not privately profitable. As a result, Conchita makes the socially inefficient decision not to put on the display.

Although the private market fails to supply the fireworks display demanded by Hereza residents, the solution to Hereza's problem is obvious: the local government can sponsor a Twelfth Night celebration. The town council raises revenue by levying taxes on property in the area (in Spain these local property taxes are known as *Impuesto sobre Bienes Inmuebles* (IBI)). Suppose that the council uses this mechanism so as to raise on average an extra €2 a year from every resident of Hereza and then uses the resulting revenue to hire Conchita to produce the fireworks. Everyone in Hereza is better off by €8 – the €10 in value from the fireworks minus the €2 tax bill. Conchita can help Hereza reach the efficient outcome as a public employee even though she could not do so as a private entrepreneur.

The story of Hereza is simplified, but it is also realistic. In fact, many local councils in Spain do pay for fireworks for festivals, as do local councils in the United Kingdom on 5 November; local governments in France pay for fireworks on 14 July (Bastille Day) and many local governments in the USA pay for fireworks on 4 July (Independence Day). Moreover, the story shows a general lesson about public goods: because public goods are not excludable, the free rider problem prevents the private market from supplying them. The government, however, can potentially remedy the problem. If the government decides that the total benefits exceed the costs, it can provide the public good and pay for it with tax revenue, making everyone better off.

Some Important Public Goods

There are many examples of public goods. Here we consider three of the most important.

National Defence The defence of the country from foreign aggressors is a classic example of a public good. Once the country is defended, it is impossible to prevent any single person from enjoying the benefit of this defence. Moreover, when one person enjoys the benefit of national defence, he does not reduce the benefit to anyone else. Thus, national defence is neither excludable nor rival.

National defence is also one of the most expensive public goods. In the UK in 2013 it accounted for about €56 billion of government expenditure – the fourth largest category (behind social security, the National Health Service and education). In France the defence budget in 2013 was around €31.4 billion, in Germany, around €31.5 billion and in the United States a massive €530 billion! People disagree about whether these amounts are too small or too large, but almost no one doubts that some government spending on national defence is necessary. Even economists who advocate small government agree that the national defence is a public good the government should provide.

Basic Research The creation of knowledge is a public good. If a mathematician proves a new theorem, the theorem enters the general pool of knowledge that anyone can use without charge. Because knowledge is a public good, profit-seeking firms tend to free ride on the knowledge created by others and, as a result, devote too few resources to creating new knowledge.

In evaluating the appropriate policy towards knowledge creation, it is important to distinguish general knowledge from specific, technological knowledge. Specific, technological knowledge, such as the invention of a better battery, can be patented. The inventor thus obtains much of the benefit of their invention, although certainly not all of it. By contrast, a mathematician cannot patent a theorem; such general knowledge is freely available to everyone. In other words, the patent system makes specific, technological knowledge excludable, whereas general knowledge is not excludable.

The government tries to provide the public good of general knowledge in various ways. Government agencies, such as the UK's Research Councils, subsidize basic research in many academic disciplines, including medicine, mathematics, science and even economics. Some people justify government funding of space exploration programmes – for example by public funding of the European Space Agency (ESA) or the US National Aeronautics and Space Administration (NASA) – on the grounds that it adds to society's

pool of knowledge. Certainly, many private goods, including bullet-proof vests and the instant drink Tang, an orange-flavoured non-carbonated drink the brand name of which is now owned by Kraft, use materials that were first developed or made popular by scientists and engineers trying to land a man on the moon. Determining the appropriate level of governmental support for these endeavours is difficult because the benefits are hard to measure. Moreover, the politicians who appropriate funds for research usually have little expertise in science and, therefore, are not in the best position to judge what lines of research will produce the largest benefits.

Fighting Poverty Many government expenditure programmes are aimed at helping the poor. These anti-poverty programmes are financed by taxes on families that are financially more successful.

Economists disagree among themselves about what role the government should play in fighting poverty. Here we note one important argument: advocates of anti-poverty programmes claim that fighting poverty is a public good.

Suppose that everyone prefers to live in a society without poverty. Even if this preference is strong and widespread, fighting poverty is not a 'good' that the private market can provide. No single individual can eliminate poverty because the problem is so large. Moreover, private charity is hard pressed to solve the problem: people who do not donate to charity can free ride on the generosity of others. In this case, taxing the wealthy to raise the living standards of the poor can make everyone better off. The poor are better off because they now enjoy a higher standard of living, and those paying the taxes are better off because they enjoy living in a society with less poverty.

The Difficult Job of Cost–Benefit Analysis

So far we have seen that the government provides public goods because the private market on its own will not produce an efficient quantity. Yet deciding that the government must play a role is only the first step. The government must then determine what kinds of public goods to provide and in what quantities.

Suppose that the government is considering a public project, such as building a new motorway or autobahn. To judge whether to build the motorway, it must compare the total benefits of all those who would use it to the costs of building and maintaining it. To make this decision, the government might hire a team of economists and engineers to conduct a study, called a **cost–benefit analysis**, the goal of which is to estimate the total costs and benefits of the project to society as a whole.

cost–benefit analysis a study that compares the costs and benefits to society of providing a public good

Cost–benefit analysis might be used when considering major infrastructure projects such as building dams, bridges, new railway lines, opening up canals and waterways for freight (and leisure) traffic, developing new port facilities, introducing or extending speed cameras on the road network, developing new water treatment and distribution networks, public transport networks such as subways and tram systems, investing in networks to improve internet access for both businesses and private consumers, extending or constructing new airports and so on.

Cost–benefit analysts have a tough job. Because a motorway will be available to everyone free of charge, there is no price with which to judge the value of the motorway. Simply asking people how much they would value the motorway is not reliable: quantifying benefits is difficult using the results from a questionnaire, and respondents have little incentive to tell the truth. Those who would use the motorway have an incentive to exaggerate the benefit they receive to get the motorway built. Those who would be harmed by the motorway have an incentive to exaggerate the costs to them to prevent the motorway from being built.

The efficient provision of public goods is intrinsically more difficult than the efficient provision of private goods. Private goods are provided in the market. Buyers of a private good reveal the value they place on it by the prices they are willing to pay when they purchase and thus use a good (this is referred to

as *revealed preference*). Sellers reveal their costs by the prices they are willing to accept. By contrast, cost–benefit analysts do not observe any price signals when evaluating whether the government should provide a public good. Their findings on the costs and benefits of public projects are, therefore, rough approximations at best.

Contingent Valuation Methods (CVM) To overcome some of the problems associated with cost–benefit analysis, contingent valuation can be used. This method is a survey–based approach which aims to place a monetary value on a good through getting respondents to state a preference and a willingness to pay. The difference between revealed preference as outlined above and stated preference is that the latter allows a valuation based on non-use, that is the value that individuals place on a good even if they do not use them. Questions in surveys provide respondents with options which attempt to discover their preferences in both how much they are prepared to give up to secure the benefit (for example, reduced pollution), termed willingness-to-pay (WTP) and how much they would need to be paid to put up with a cost (for example, a certain level of pollution) termed willingness to accept (WTA). For example, CVM might be used when seeking to find out what value individuals place on having improved water and sewage treatment services.

In preparing the survey, care needs to be taken to ensure that the good itself and the proposed changes to the good are clearly defined so that respondents can more easily understand what they are being asked to value. The data collected allow researchers to be able to link WTP and WTA responses to changes in utility measured in monetary units. However, CVM presents difficulties in that respondents may not be familiar with the scenarios and the choices they are being asked to make. In addition, it is not always clear how seriously respondents treat the questions or think about their responses because there is no cost associated to them in doing so. As a result the reliability of the surveys can be questioned. One support for such a criticism is the differences observed between WTP and WTA, which, in theory, ought to be very similar if not the same since the respondent is being asked to value opposites.

CASE STUDY How Much is a Life Worth?

Ask yourself this question: how much would you say a human life is worth? Or, put it another way, how much extra, per year, would you be prepared to pay in taxes to save one additional human life? €10, €10,000, €1,000,000? The question is almost impossible to answer, in part because it is so subjective (a normative proposition) and in part because most of us do not have €1,000,000 or anything like that to make the question feasible.

Imagine that you have been elected to serve as a member of your local town council. The town engineer comes to you with a proposal: the town can spend €10,000 to build and operate traffic lights at the junction of Main Street (the town's busy main road) and Easy Street (the much quieter, leafy road where the stockbrokers live). At present, there are only signs instructing motorists entering Main Street from Easy Street that they must stop and give way to Main Street traffic. The benefit of the traffic lights is increased safety. The engineer estimates, based on data from similar intersections, that the traffic lights would reduce the risk of a fatal traffic accident over the lifetime of the traffic lights from 1.6 to 1.1 per cent. Should you spend the money for the new lights?

To answer this question, you turn to cost–benefit analysis. But you quickly run into an obstacle: the costs and benefits must be measured in the same units if you are to compare them meaningfully. The cost is measured in euros, but the benefit – the possibility of saving a person's life – is not directly monetary. To make your decision, you have to put a monetary value on a human life.

At first, you may be tempted to conclude that a human life is priceless. After all, there is probably no amount of money that you could be paid to voluntarily give up your life or that of a loved one. This suggests that a human life has an infinite monetary value.

For the purposes of cost–benefit analysis, however, this answer leads to nonsensical results. If we truly placed an infinite value on human life, we should be placing traffic lights on every street corner. Similarly, we should all be driving large cars with all the latest safety features, instead of smaller ones

with fewer safety features. Yet traffic lights are not at every corner, and people sometimes choose to buy small cars without side-impact air bags or antilock brakes. In both our public and private decisions, we are at times willing to risk our lives to save some money.

Once we have accepted the idea that a person's life does have an implicit monetary value, how can we determine what that value is? One approach, sometimes used by courts to award damages in wrongful death suits, is to look at the total amount of money a person would have earned if he or she had lived. Economists are often critical of this approach. It has the bizarre implication that the life of a retired or disabled person has no value. However, the earnings potential idea has been used as one way of measuring the value of a human life.

Health services across Europe have to find some way of making decisions about who to treat and who not to treat because the demand for health services far outstrips supply. The basis for such a decision can be to put an estimate on the contribution the individual might make to society based on their future earning capacity. Take the example of a patient who is 45. In theory that person has another 20 years of work left in him before retirement (if the official retirement age is 65 – but this is changing in many countries). If the average wage is €25,000 per annum, that means that this individual is worth €500,000 to society. If the treatment, therefore, costs less than €500,000 it could be argued it is worth treating this individual. One method that has been designed to try to come to terms with this is a unit of measurement of the outcome of treatment called quality-adjusted life years (QALYs). The concept was developed by health economics researchers at the University of York in the UK. A QALY gives a measure of the extra quality of life and the number of years of extra life resulting from a treatment.

To create a scale (the measure of health state) that covers the range of outcomes of a treatment (or intervention as it is called), researchers went from 1, the best possible state of health where all normal functions were at the disposal of the patient, to 0, the worst possible health state. Between these two scales (and it may seem rather bizarre) was 'death' – in other words, an intervention could be carried out that results in the death of the patient.

What could be worse than death? This is very much a normative issue but it may be possible to imagine a state of health that exists where a patient suffers intolerable physical and mental pain. Doctors can cite cases where patients are in a state where quality of life is both terrible and permanent and the patient has no feelings other than continuing pain. In such tragic situations, some might view the quality of life as being worse than being dead. Indeed, cases where patients in this state argue for the right to have their life terminated is one example where this situation might exist.

The researchers drew up the grades within this scale based on responses from patients about their preferences for different states of health. The grades within the scale, therefore, allow the user to be able to calculate the number of QALYs gained from a treatment. If treatment of a condition of a 50-year-old patient restores the person to full health and that person's life expectancy is 76, then the treatment will have resulted in 26 QALYs. If the treatment restores the person to a reasonable state of health but with some disabilities, such as some limitations to walking, breathing, sight, hearing or other problems rated on the scale as 0.75, then the number of QALYs would be $0.75 \times 26 = 19.5$ QALYs.

Having got some indication of an outcome, the cost of providing that outcome can be identified and as a result some measure of the value for money provided by health care treatment can be identified. Assume that the cost of three treatments is as follows:

- The average cost of a primary hip replacement is €4,160.
- The average cost of a heart transplant is €16,500.
- The average cost of treating a patient with an open lower limb fracture or dislocation is €2,521 per fracture.

Let us assume that the outcomes outlined in Table 10.1 are possible after treating a 40-year-old patient.

Predictably the fracture comes out as providing the best 'value for money' because of the amount of years of 'good health' it provides after treatment. We can also deduce from this that the opportunity cost of treating one person with a heart transplant is 6.8 people that could have been given a hip replacement and achieving the same cost per QALY – in other words, the cost per outcome.

TABLE 10.1

| Treatment | Cost per patient (€) | Measure of health state | Additional life expectancy | QALY | Cost per QALY (€) |
|------------------|----------------------|-------------------------|----------------------------|-------|-------------------|
| Hip replacement | 4,160 | 0.95 | 15 | 14.25 | 291.93 |
| Heart transplant | 16,500 | 0.83 | 10 | 8.3 | 1,987.95 |
| Fracture | 2,521 | 1.00 | 35 | 35 | 72.03 |

This system therefore provides the health professionals with some form of measure of the outcome of treatment, its associated cost, comparisons between hospitals and departments, and thus the basis for decision making.

Another way to value human life is to look at the risks that people are voluntarily willing to take and how much they must be paid for taking them. Mortality risk varies across jobs, for example. Construction workers in high-rise buildings face greater risk of death on the job than office workers do. By comparing wages in risky and less risky occupations, controlling for education, experience and other determinants of wages, economists can get some sense about what value people put on their own lives. Studies using this approach conclude that the value of a human life is about €10 million.

Other models look at the individual's willingness to pay to buy an extra year's worth of life or how much someone would be willing to pay to reduce the risk of death. There are inherent problems with all of these models but the essential basis of each of them is the fundamental concept of a trade-off as outlined in Principle 1 of the *Ten Principles of Economics*.

One economist, Orley Ashenfelter, calculated the value of a human life based on an analysis of the effect of a rise in speed limits on US roads. The rise in the speed limit meant that drivers drove slightly faster. This 'led to' a rise in deaths on the roads by 35 per cent. Ashenfelter then compared this to the data from areas where speed limits had not been changed. People in such areas drove slower.

The trade-off occurs between the numbers of hours saved as a result of faster travel against the lives lost as a result. Ashenfelter found that the saving in time was about 45 million hours, whereas the increase in the number of deaths was 360. Dividing the two gives an average of 125,000 hours per life. Multiply this figure by the US average wage and he ended up with the figure of \$980,000.

Accidents will happen but how much extra would individuals be willing to pay, for example in higher prices for train travel, to reduce the risk of accidents to almost zero?



One problem with such calculations is that it might assume that everyone's life is of equal value. Should the value of the life of a 98-year-old be valued differently to that of a 25-year-old? The question then arises, where do you draw the line? When does someone become too old to treat in a hospital? Already there have been cases around Europe where doctors have refused to treat patients who have smoked and drank heavily despite repeated warnings because they feel they have brought the problems onto themselves and that the use of public money to treat such people is misdirected.

The basis of such decision making, therefore, is to find some way in which human life can be valued and then set the benefits (measured by some agreed monetary value) against the costs of achieving the outcome, whether this be receiving hospital treatment or investing in measures to improve safety on the roads or other forms of transport and in the workplace. If the benefits outweigh the cost, this provides the basis for making the decision. We can now return to our original example and respond to the town engineer. The traffic light reduces the risk of fatality by 0.5 percentage points. Thus, the expected benefit from having traffic lights is $0.005 \times \text{€}10 \text{ million}$, or €50,000. This estimate of the benefit well exceeds the cost of €10,000, so you should approve the project.

SELF TEST What is the free rider problem? • Why does the free rider problem induce the government to provide public goods? • How should the government decide whether to provide a public good?

The Optimal Provision of a Public Good

Governments provide public goods because they provide a benefit to society as a whole. We can assume, therefore, that governments should continue to provide a public good up to a point where the marginal benefit gained from an extra unit provided is equal to the marginal cost of providing that extra unit. When considering the marginal cost of providing public goods we also have to consider the opportunity cost of the resources used in the provision of public goods. The marginal benefit of an extra unit of provision of a public good will be shared by a large number of consumers because of the fact that public goods are non-rival. To find the total benefit of the provision of public goods we have to add up all the marginal benefits that users gain.

This is represented in Figure 10.2. On the vertical axis is price in euros and on the horizontal axis the quantity of the public good provided. In reality there are large numbers of people who benefit from public goods; in Figure 10.2 we assume there are two consumers denoted by the demand curves D_1 and D_2 . We could extend the number of demand curves to match the number of people in society but providing the simplification of representing society with just two people illustrates the principle just as well.

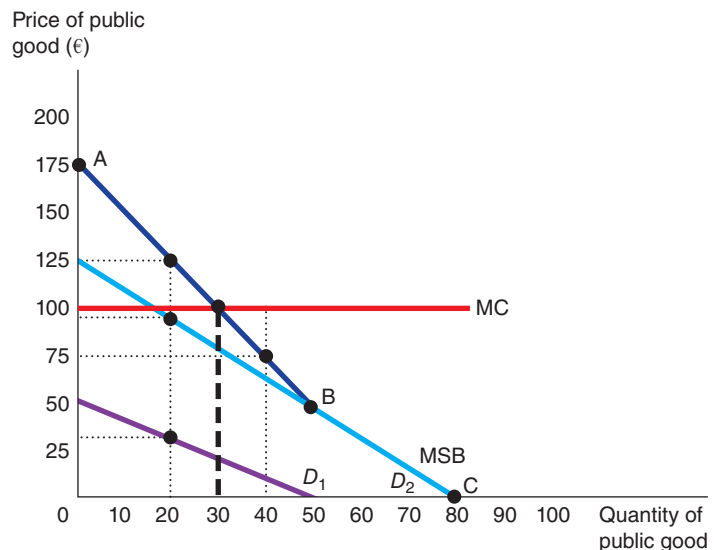
The demand curve D_1 represents the first consumer's demand for the public good and demand curve D_2 the second consumer's demand for the public good. Demand curve D_1 shows that consumer 1 places a value on the 20th unit of the public good provided of €31.25 and demand curve D_2 shows that consumer 2 places a value on the 20th unit of the public good provided of €93.75. Because the good is non-rival, both gain benefits from using the public good and so the sum of the benefits of the two consumers of the 20th unit provided is the vertical summation of these two values (€125). We can sum the benefits of both consumers at quantities of the public good from 0 to 80 units and we get the marginal social benefit (MSB) curve which is a kinked demand curve shown as ABC which is the vertical summation of the two individual consumer demand curves.

The optimum provision of this public good would be where the marginal cost of providing the public good intersects the MSB curve. If the marginal cost of providing this public good were €100 per unit then the optimum provision would be 30 units of the public good. If the government provided more of this public good (say 40 units), the marginal cost of providing this amount would exceed the MSB which at 40 units is €75. It would also not be efficient to produce anything less than 30 units (20 for example), because the MSB at 20 units (€125) is greater than the marginal cost of provision (€100) so it would be appropriate to expand provision of this public good.

FIGURE 10.2

The Optimum Provision of a Public Good

Society's collective benefit from the provision of a public good is given by the two demand curves D_1 and D_2 . The sum of the benefits of these two consumers is the vertical summation of the value each places on the marginal unit of the public good supplied. The kinked demand curve, ABC, is the marginal social benefit (MSB) curve. The optimum provision of this public good is the intersection of the marginal cost of providing this public good with the MSB curve when 30 units are provided and the MSB at this level of provision is equal to the marginal cost at €100.



As we have noted many times so far, this is a model to illustrate a point. In this case the point is that in deciding on an appropriate allocation of resources for the provision of public goods, some consideration has to be given to the costs of providing that good and the benefit accruing to society as a whole of the provision. Economists will try to quantify the costs and benefits and whilst it will never be possible to fully quantify such costs and benefits, an attempt to do so enables a more informed decision to be made which might better meet society's needs and use scarce resources more effectively.

COMMON RESOURCES

Common resources, like public goods, are not excludable: they are available free of charge to anyone who wants to use them. Common resources are, however, rival: one person's use of the common resource reduces other people's ability to use it. Thus, common resources give rise to a new problem. Once the good is provided, policymakers need to be concerned about how much it is used. This problem is best understood from the classic parable called the **Tragedy of the Commons**.

Tragedy of the Commons a parable that illustrates why common resources get used more than is desirable from the standpoint of society as a whole

The Tragedy of the Commons

Consider life in a small town and of the many economic activities that take place in the town. One of these activities is raising sheep. Many of the town's families own flocks of sheep and support themselves by selling the sheep's wool, which is used to make clothing.

As our story begins, the sheep spend much of their time grazing on the land surrounding the town, called the Town Common. No family owns the land. Instead, the town residents own the land collectively, and all the residents are allowed to graze their sheep on it. Collective ownership works well because land is plentiful. As long as everyone can get all the good grazing land they want, the Town Common is not a rival good, and allowing residents' sheep to graze for free causes no problems. Everyone in town is happy.

As the years pass, the population of the town grows, and so does the number of sheep grazing on the Town Common. With a growing number of sheep and a fixed amount of land, the land starts to lose its ability to replenish itself. Eventually, the land is grazed so heavily that it becomes barren. With no grass left on the Town Common, raising sheep is impossible, and the town's once prosperous wool industry disappears and, tragically, many families lose their source of livelihood.

What causes the tragedy? Why do the shepherds allow the sheep population to grow so large that it destroys the Town Common? The reason is that social and private incentives differ. Avoiding the destruction of the grazing land depends on the collective action of the shepherds. If the shepherds acted together, they could reduce the sheep population to a size that the Town Common could support. Yet no single family has an incentive to reduce the size of its own flock because each flock represents only a small part of the problem.

In essence, the Tragedy of the Commons arises because of an externality. When one family's flock grazes on the common land, it reduces the quality of the land available for other families. This is a social cost. Because people neglect the social cost when deciding how many sheep to own, the result is an excessive number of sheep.

If the tragedy had been foreseen, the town could have solved the problem in various ways. It could have regulated the number of sheep in each family's flock, taken account of the social cost by taxing sheep, or auctioned off a limited number of sheep-grazing permits.

In the case of land, however, there is a simpler solution. The town can divide up the land among town families. Each family can enclose its allotment of land with a fence and then protect it from excessive

grazing. In this way, the land becomes a private good rather than a common resource. This outcome occurred during the enclosure movement in England in the 17th century.

The Tragedy of the Commons is a story with a general lesson: when one person uses a common resource, he diminishes other people's enjoyment of it. Because of this negative externality, common resources tend to be used excessively. The government can solve the problem by reducing use of the common resource through regulation or taxes. Alternatively, the government can sometimes turn the common resource into a private good.

This lesson has been known for thousands of years. The ancient Greek philosopher Aristotle pointed out the problem with common resources: 'What is common to many is taken least care of, for all men have greater regard for what is their own than for what they possess in common with others.'

Some Important Common Resources

There are many examples of common resources. In almost all cases, the same problem arises as in the Tragedy of the Commons: private decision makers use the common resource too much. Governments often regulate behaviour or impose fees to mitigate the problem of overuse.

Clean Air and Water Markets do not adequately protect the environment. Pollution is a social cost that can be remedied with regulations or with taxes on polluting activities (as we shall see in more detail in Chapter 11). One can view this market failure as an example of a common-resource problem. Clean air and clean water are common resources like open grazing land, and excessive pollution is like excessive grazing. Environmental degradation is a modern Tragedy of the Commons.

Congested Roads Roads can be either public goods or common resources. If a road is not congested, then one person's use does not affect anyone else. In this case, use is not rival, and the road is a public good. Yet if a road is congested, then use of that road yields a social cost. When one person drives on the road, it becomes more crowded, and other people must drive more slowly. In this case, the road is a common resource.

One way for the government to address the problem of road congestion is to charge drivers a toll. A toll is, in essence, a tax on the social cost of congestion. Often, as in the case of local roads, tolls are not a practical solution because the cost of collecting them is too high. Nevertheless, tolls are often charged on stretches of motorways in continental Europe and the USA, and occasionally in the UK (on the M6, for example).

Sometimes congestion is a problem only at certain times of day. If a bridge is heavily travelled only during rush hour, for instance, the congestion social cost is larger during this time than during other times of day. The efficient way to deal with these social costs is to charge higher tolls during rush hour. This toll would provide an incentive for drivers to alter their schedules and would reduce traffic when congestion is greatest.

Another policy that responds to the problem of road congestion, is a tax on petrol. Petrol is a complementary good to driving: an increase in the price of petrol tends to reduce the quantity of driving demanded. Therefore, a petrol tax reduces road congestion. A petrol tax, however, is an imperfect solution to road congestion. The problem is that the petrol tax affects other decisions besides the amount of driving on congested roads. For example, the petrol tax discourages driving on non-congested roads, even though there is no congestion social cost for these roads.

Fish, Whales and Other Wildlife Many species of animals are common resources. Fish and whales, for instance, have commercial value, and anyone can go to the ocean and catch whatever is available. Each person has little incentive to maintain the species for the next year. Just as excessive grazing can destroy the Town Common, excessive fishing and whaling can destroy commercially valuable marine populations.

The ocean remains one of the least regulated common resources. Two problems prevent an easy solution. First, many countries have access to the oceans, so any solution would require international

cooperation among countries that hold different values. Second, because the oceans are so vast, enforcing any agreement is difficult. As a result, fishing rights have been a frequent source of international tension among normally friendly countries.

Within the United Kingdom and other European countries, various laws aim to protect fish and other wildlife. For example, the government charges for fishing and hunting licences, and it restricts the lengths of the fishing and hunting seasons. Fishermen are often required to throw back small fish, and hunters can kill only a limited number of animals or shoot certain wild birds such as pheasant or grouse during specified periods of the year. All these laws reduce the use of a common resource and help maintain animal populations.

SELF TEST Why do governments try to limit the use of common resources?

MERIT GOODS

Some goods can be provided by the market mechanism but if left purely to the market they may be under-consumed. These types of goods are termed **merit goods**. Merit goods arise because consumers may have imperfect information about the benefits of these goods and are not able to value them appropriately as a result. If the market signal does not fully convey the value of the benefits to consumers then they are likely to under-invest in these goods. The benefits to consumers may occur sometime in the future but the price the consumer is being asked to pay occurs in the present. Merit goods are an example of an **intertemporal choice** problem in economics – the term intertemporal relates to decision making over time when current decisions affect choices made in the future.

merit goods goods which can be provided by the market but may be under-consumed as a result
intertemporal choice where decisions made today can affect choices facing individuals in the future

Examples of merit goods include education, health care, pensions and insurance. In each of these cases the market can provide these goods; there are plenty of examples of private firms providing education (schools and universities run by private firms on the basis of students paying fees to attend); private health care through firms such as BUPA, Spire, BMI Healthcare and Capio in Sweden may be profit or non-profit organizations, but access to services is through payment at the point of use, unlike publicly provided health care services which are free at the point of use and financed by the state through taxpayers primarily. Equally, there are plenty of firms providing a range of insurance and pension services who are part of the private sector.

Education as a Merit Good

If you are reading this then there is a very good chance that you are at university. You will have made a decision to invest in your future by pursuing an undergraduate degree. In so doing you are sacrificing earnings that you could have generated and may also be incurring a considerable debt burden which you will have to pay off at some point in the future. One of the reasons why you have chosen to study is likely to be that you hope getting a degree will enable you to get a job which you will enjoy and which will also give you a higher salary. The lifetime earnings of those with a degree are higher than those without.

However, what you do not know at this point in time is the precise nature of the future benefits you will gain or how your degree will enable you to make different choices in the future, choices which may bring you considerable benefits which you simply cannot calculate at this point in time. Because of this

imperfect information, some people may choose not to go to university. There are not only private benefits to the individual of going to university, there are also social benefits. A better educated workforce is more likely to be more productive and if society's stock of human capital is more productive then standards of living overall tend to be higher. If you get a better job then you will pay more taxes and so government will be able to provide more services and this also benefits society. When looking at issues surrounding the commons we noted that there were social costs associated with consumption, costs that affect people other than the decision maker. In the case of a merit good, e.g. education, the problem is that there are social benefits which are not considered by the individual in their decision making.

When looking at school education the issue is a little more complicated. The decision to send a child to school is invariably the parents, but the beneficiary is the child. This is a classic principal-agent problem. The principal (the parent) is acting on behalf of the agent (the child) and if schooling was left to the market there might be a conflict of interest. If the parent has to pay for education but is not getting the benefit then there is an incentive for the parent to not send their child to school which leads to under-consumption of education.

Training of workers is another aspect of education that is under-consumed. Firms often complain that the mainstream education system does not adequately prepare young people for the world of work. When an employee gets a job, firms invariably have to spend money training them but the amount spent on training is likely to be far less than would be the socially optimum level. Spending on training would yield some private benefit to the firm in the form of improved productivity but there would also be some social benefit of this improved productivity as noted above. The worker would also be in a stronger position to access new employment if she were made redundant, thus reducing the potential costs to the state of having to provide benefits.

The firm may not invest the full amount on training despite the private benefit because it fears that it might not receive the full benefit of the investment. Workers might be better skilled as a result and they may find work with rival firms who gain the benefit of the investment in training that the original firm made.

It is for these reasons that the state provides education and training for people across Europe and subsidizes education and training in the workplace.

Health Care, Insurance and Pensions as Merit Goods

Few people are in a position to be able to judge when and if they need health care and insurance. For many young people, the prospect of saving today to fund a pension in 30–40 years' time is a decision they do not feel in a position to make – the benefits are too far in the future to be meaningful. If individuals had to pay for their own health care, insurance and pension provision then there would be an incentive to under-consume.

Many young people, for example, feel healthy and are in work and find it hard to conceive of what it might be like to be ill or what income they might need to live when they retire. The price of acquiring sufficient health care cover or of securing a sufficient pension provision is likely to be seen as too high and so some would not spend money on this necessary provision.

Some people look at insurance in the same way. Paying for life assurance and insuring home and contents is a gamble – we all know that our lives will end at some point but we don't know when, and when it does happen it will not be us that get the benefit of life cover, it will be someone else. We all know that there is a risk that our homes could be burgled or damaged by fire or some other disaster but it might never happen throughout our lives. There is an incentive, therefore, to put off taking out life cover or insurance on our homes. When we do die we might be leaving our loved ones with little or no income and/or the necessity of paying off mortgages or other debt which could prove catastrophic. If our homes are subject to some disaster or crime then the cost of putting our property back to a state it was in before the incident could be significant. The problem is that, in each case, individuals do not have the information to make informed (rational) decisions on the relative costs and benefits and so these goods are under-consumed as a result.

The state can intervene in these cases to force firms and individuals to contribute to a pension scheme, to provide health care free at the point of use via taxes and to force people to take out insurance (most countries make taking out vehicle insurance, for example, a legal requirement of owning a vehicle). In these cases, it can be argued that individuals are not in the best place to be able to judge the benefits of consumption (both private and social) and so the state intervenes to ensure that goods which have merit are provided.

De-merit Goods

We have made reference to the distinction between 'goods' and 'bads'. Not all products which are consumed are good for us. **De-merit goods** have the characteristic of being over-consumed if left to the market. This is because consumption of these goods imposes private and social costs and, in making a consumption decision, the individual does not have the information to fully understand these costs. Tobacco, alcohol, pornography and non-medicinal drugs are examples of de-merit goods.

de-merit goods goods that are over-consumed if left to the market mechanism and which generate both private and social costs which are not taken into account by the decision maker

When individuals consume goods like tobacco and alcohol, even though they are legal in many countries, there are private costs to the individual of the consumption. The damage to health that both of these goods generate may be partly known by the consumer but the full extent of the damage they are causing themselves is unknown. In addition, consumption of these goods might be associated with addiction and as a result it becomes harder for the individual to break the habit of smoking and drinking.

In addition to the private costs there are also social costs. The cost to the health care system of treating patients with tobacco and alcohol related illness and disease is extensive and the resources spent on this care could be used to treat other patients and conditions – in other words, the decision of smokers and drinkers diverts resources from what might be argued to be more socially efficient allocations. In addition, alcohol consumption is cited as a cause of anti-social behaviour and the resources needed to deal with these problems represent a cost to society as a whole and also diverts these resources away from other uses. The police, for example, could be dealing with other types of crime but if they have to devote resources to policing town and city centres into the night to deal with drunken groups of people, and the violence and crime that can arise as a result of excessive alcohol consumption, they cannot deal with other types of crime.

There is a case, therefore, for the state to intervene to regulate or legislate these markets, to tax products deemed to have private and social costs to reduce consumption.

CONCLUSION

In this chapter, we have seen there are some 'goods' that the market does not provide adequately. Markets do not ensure that the air we breathe is clean or that our country is defended from foreign aggressors. Instead, societies rely on the government to protect the environment and to provide for the national defence and to help provide goods that if left to the market would be either under or over-consumed.

There is considerable debate over the extent to which the state should get involved in the provision of public and merit goods and whether such intervention represents both a more efficient and equitable allocation of scarce resources. These debates are at the heart of issues which are not only influenced by economic understanding but also politics. Politics is about power and it may be that different groups in society have the power to influence decision making more extensively than others and so resource allocation can be further distorted. The issues which have been touched upon in this chapter will be revisited on numerous occasions during the course of your studies and we have attempted to introduce some of the basic principles that underpin these issues.

IN THE NEWS



The Problem of the Commons: Still Unsettled after 100 Years

As this chapter has shown, awareness about the Tragedy of the Commons is widespread but that does not mean that we have solved all the problems that the commons presents. Indeed, it can be argued that there are still considerable problems that exist to which humans are still searching for solutions.

In an article in the *American Economic Review* in 2011, Robert N. Stavins, Albert Pratt Professor of Business and Government at Harvard Kennedy Business School, explored some of these issues. Stavins notes the concerns over air and water quality, hazardous waste, species extinction, the maintenance of stratospheric ozone and the stability of the global climate as being examples of cases where common property issues exist over which no overarching authority can offer complete control.

Stavins deals with categories of common problems which include natural resources, environmental quality and the global climate. Whilst there have been concerns expressed over the depletion of non-renewable resources such as minerals and fossil fuels, Stavins notes that price signals have meant that these resources are not becoming more scarce. In contrast, renewable resources such as forests and fisheries have become more scarce to the extent that some have become exhausted and species extinction has occurred. One of the main reasons for this is that non-renewable resources are invariably owned by someone and as a result there is a regulation of the use of these resources by the market. With commons such as

forests and fisheries, such ownership is less clear and the consequences as explained in the Tragedy of the Commons is more likely.

Why have renewable resources been depleted to this extent? This can be explained through the use of the concept of scarcity rent. Scarcity rent refers to the marginal opportunity cost imposed on future generations from the extraction of one more unit of a resource such as fish from the sea. Fishermen will take into consideration the marginal revenue they receive from catching fish in relation to the marginal extraction cost they incur. What they do not take into account is scarcity rent. All fisherman, therefore, are carrying out their activity and not taking into consideration the value of the asset which they are exploiting. Each is incentivized to maximize their catch but not incur the full cost of the reduction in the value of the asset – the fish stocks in the sea.

Examples where the difference between estimated efficient extraction data and actual data provided by empirical studies confirm the theory. The efficient number of ships fishing for halibut in the Bering Seas is estimated at nine but there were over 140. In New England, the efficient number of lobster traps was put at 450,000 but there were in excess of 1 million used. When this happens there are too many resources devoted to fishing and as fish stocks are reduced the future profitability is also reduced.

When considering the issue of environmental quality the issues are slightly different. The commons problem here is that the air, water and land are being used

as a repository for pollution and waste. This creates social costs and occurs because ownership of the air, land and water is insufficiently clear. Economists accept that it is not possible to reduce this pollution to zero but instead the focus is on what is cost effective – minimizing the cost of achieving some desired objective which might be seen as being acceptable levels of pollution. Crucial to this cost-effectiveness calculation is a concept called the marginal abatement cost (MAC).

Abatement refers to the reduction in pollution and the MAC is the cost expressed in terms of the last unit of pollution not emitted (abated). As firms attempt to reduce pollution the MAC is likely to rise as it becomes harder to develop the technology and processes to cut pollution. This implies that firms who are the heaviest polluters initially will be able to contribute most to pollution reduction schemes because the MAC to them is relatively low initially. As pollution reduction levels increase, the MAC rises and there will come a point when it will not be commercially feasible for firms to reduce pollution any further.

If standards are set which apply to all firms in respect of pollution reduction or pollution emission control, monitoring and forcing such standards can be costly because firms will differ in the technology they employ, their production processes and the age and efficiency of their assets (which might contribute to how much pollution they emit in the first place), among other things, and hence have very different MACs.

The much reported issue of greenhouse gas emissions and the

(Continued)

estimated effects on the global climate and its subsequent effects on humans and the environment is what Stavins refers to as the ultimate commons problem. Climate change is a commons problem because greenhouse gases remain in the atmosphere for significant periods of time. Once gases have been emitted from one country they mix in the atmosphere and any subsequent cost could be imposed on another country or countries regardless of whether they have been the cause of the problem in the first place. Indeed, the argument about who is responsible has been at the centre of disputes between nations at climate change summits. Developed nations point to the increase in emissions coming from emerging countries both now and in the future if they continue to grow at rates seen in recent years. The emerging

nations point to the fact that much of the emissions that are estimated to cause climate problems has been caused by the developed nations over the last 100 years.

Therefore, what incentive is there for any individual nation to take steps to reduce greenhouse gas emissions because the benefits to them individually will be far less than the costs and as a result a free-rider problem exists. To overcome this, a global level of cooperation is necessary and, as the outcomes of recent summits has shown, this is not easy to achieve.

Solutions to these commons problems may lie in taxes on carbon emissions, which increase the cost of production and provide an incentive to reduce emissions and on using the market mechanism to price emissions. Governments or some international body sets a cap

on the level of emissions deemed appropriate and firms buy permits to allow them to emit that level. If they can cut emissions they can sell the excess allowance on the open market. Firms who are unable to achieve the cap set, for whatever reason, can buy these spare allowances. The incentive for firms is to reduce emissions and if the cap keeps being reduced emissions levels can begin to be cut.

We will look at these schemes in more detail in later chapters.

Questions

- 1 Why can water quality, hazardous waste, species extinction, the ozone layer and the stability of the global climate be seen as examples of the commons?
- 2 Explain why non-renewable resources might not be subject to increased scarcity but renewable resources might be.
- 3 Why does a lack of consideration of scarcity rent by fishermen mean that stocks of fish are likely to be depleted to dangerous levels?
- 4 Why do you think marginal abatement costs are likely to be relatively low at first but then rise sharply as abatement increases?
- 5 Why does the temporal and spatial characteristics of greenhouse gas emissions present problems to policymakers in seeking to reduce emissions and why might there exist a free rider problem?

The dumping of potentially hazardous waste might suggest market signals are not working effectively leading to external costs being imposed on society as a whole.



SUMMARY

- Goods differ in whether they are excludable and whether they are rival. A good is excludable if it is possible to prevent someone from using it. A good is rival if one person's use of the good reduces other people's

ability to use the same unit of the good. It can be argued that markets work best for private goods, which are both excludable and rival. Markets do not work as well for other types of goods.

- Public goods are neither rival nor excludable. Examples of public goods include fireworks displays, national defence and the creation of fundamental knowledge. Because people are not charged for their use of the public good, they have an incentive to free ride when the good is provided privately. Therefore, governments provide public goods, making their decision about the quantity based on cost–benefit analysis.
- Common resources are rival but not excludable. Examples include common grazing land, clean air and congested roads. Because people are not charged for their use of common resources, they tend to use them excessively. Therefore, governments try to limit the use of common resources.
- Merit goods such as education and health might be under-consumed if left to the market and so the state can step in to help provide services which provide social as well as private benefits.
- De-merit goods are goods which are over-consumed and which confer both private and social costs. Governments might intervene in the market to reduce consumption in some way either through the price mechanism (levying taxes on these goods, for example), or through regulation and legislation.

QUESTIONS FOR REVIEW

- 1 What is the difference between the public sector and the private sector? Give an example of goods and services provided by each.
- 2 Explain what is meant by a good being ‘excludable’. Give three examples of goods which exhibit the characteristics of being excludable.
- 3 Explain what is meant by a good being ‘rival’. Give three examples of goods which exhibit the characteristics of being rival in consumption.
- 4 Define and give an example of a public good. Can the private market provide this good on its own? Explain.
- 5 Explain how public goods might also lead to the free-rider problem. Give an example to support your answer.
- 6 What is cost–benefit analysis of public goods? Why is it important? Why is it hard?
- 7 Define and give an example of a common resource. Without government intervention, will people use this good too much or too little? Why?
- 8 What is the social marginal benefit curve? How can the optimum provision of a public good be calculated?
- 9 Why are merit goods under-consumed?
- 10 How might governments prevent over-consumption of de-merit goods?

PROBLEMS AND APPLICATIONS

- 1 The text says that both public goods and common resources involve social effects.
 - a. Are the social effects associated with public goods generally positive or negative? Use examples in your answer. Is the free market quantity of public goods generally greater or less than the efficient quantity?
 - b. Are the social effects associated with common resources generally positive or negative? Use examples in your answer. Is the free market use of common resources generally greater or less than the efficient use?
- 2 Think about the goods and services provided by your local government.
 - a. Using the classification in Figure 10.1, explain what category each of the following goods falls into:
 - police protection
 - road gritting
 - street lighting
 - education
 - radio broadcasts
 - rural roads
 - city streets.
 - b. Why do you think the government provides items that are not public goods?
- 3 In the UK owners of TV sets are required by law to buy a licence. Alex is a student at university and loves watching live sport on TV. He uses digital access to watch live sport, but he has not bought a TV licence.
 - a. What name do economists have for Alex?
 - b. How can the government solve the problem caused by people like Alex?
 - c. Can you think of ways the private market can solve this problem?
- 4 The text states that private firms will not undertake the efficient amount of basic scientific research.
 - a. Explain why this is so. In your answer, classify basic research in one of the categories shown in Figure 10.1.

- b. What sort of policy has the United Kingdom adopted in response to this problem?
- c. It is often argued that this policy increases the technological capability of British producers relative to that of foreign firms. Is this argument consistent with your classification of basic research in part (a)? (Hint: can excludability apply to some potential beneficiaries of a public good and not others?)
- 5 Why is there litter along most major roads but rarely in people's gardens?
- 6 An *Economist* article (19 March 1994) states: 'In the past decade, most of the rich world's fisheries have been exploited to the point of near-exhaustion.' The article continues with an analysis of the problem and a discussion of possible private and government solutions.
- a. 'Do not blame fishermen for overfishing. They are behaving rationally, as they have always done.' In what sense is 'overfishing' rational for fishermen?
- b. 'A community, held together by ties of obligation and mutual self-interest, can manage a common resource on its own.' Explain how such management can work in principle, and what obstacles it faces in the real world.
- c. 'Until 1976 most world fish stocks were open to all comers, making conservation almost impossible. Then an international agreement extended some aspects of [national] jurisdiction from 12 to 200 miles offshore.' Discuss how and why this agreement reduces the scope of the problem.
- d. The *Economist* article notes that many governments come to the aid of suffering fishermen in ways that encourage increased fishing. How do such policies encourage a vicious cycle of overfishing?
- e. 'Only when fishermen believe they are assured a long-term and exclusive right to a fishery are they likely to manage it in the same far-sighted way as good farmers manage their land.' Defend this statement.
- f. What other policies to reduce overfishing might be considered?
- 7 The demand curve for a public park for two consumers who represent society is given by:
- $$P = 150 - Q \text{ and } P = 250 - Q$$
- Graph the two demand curves and show the marginal social benefit curve for this public park. If the marginal cost of providing the park was €240, what would the optimum provision of this park be? Explain why any quantity above or below this amount would represent a less than efficient allocation.
- 8 In a market economy, information about the quality or function of goods and services is a valuable good in its own right. How does the private market provide this information? Can you think of any way in which the government plays a role in providing this information?
- 9 Do you think the Internet is a public good? Why or why not?
- 10 High-income people are willing to pay more than lower income people to avoid the risk of death. For example, they are more likely to pay for safety features on cars. Do you think cost-benefit analysts should take this fact into account when evaluating public projects? Consider, for instance, a rich town and a poor town, both of which are considering the installation of traffic lights. Should the rich town use a higher monetary value for a human life in making this decision? Why or why not?

11

EXTERNALITIES AND MARKET FAILURE

EXTERNALITIES

Firms that make and sell paper also create, as a by-product of the manufacturing process, a chemical called dioxin. Scientists believe that once dioxin enters the environment it raises the population's risk of cancer, birth defects and other health problems.

Is the production and release of dioxin a problem for society? We examined how markets allocate scarce resources with the forces of supply and demand, and we saw that the equilibrium of supply and demand is typically an efficient allocation of resources. To use Adam Smith's famous metaphor, the 'invisible hand' of the marketplace leads self-interested buyers and sellers in a market to maximize the total benefit that society derives from that market. Should we conclude, therefore, that the invisible hand prevents firms in the paper market from emitting too much dioxin?

Markets do many things well, but they do not do everything well. We saw in Chapter 1 that markets can sometimes fail and when this happens resources are not allocated to their most efficient use or the allocation is in some way considered inequitable. In this chapter we examine why markets sometimes fail to allocate resources efficiently or equitably, how government policies can potentially improve the market's allocation and what kinds of policies are likely to work best.

The market failures examined in this chapter fall under a general category called externalities. An **externality** arises when a person engages in an activity that influences the well-being of a bystander (a third party) who neither pays nor receives any compensation for that effect. If the impact on the bystander is adverse, it is called a **negative externality**; if it is beneficial, it is called a **positive externality**. Negative and positive externalities are linked to the social costs and social benefits that exist when a decision is made. Many individuals and firms make decisions based on the private costs and benefits they incur but do not always consider the social costs and benefits of their decision. As a result the price mechanism does not reflect the true cost and benefit of a decision and this can lead to a market outcome where the quantity might be privately efficient but socially inefficient and as a result the market allocation might be too high or too low. Price does not act as a true signal to consumers and producers to enable them to make informed decisions.

externality the cost or benefit of one person's decision on the well-being of a bystander (a third party) which the decision maker does not take into account when making the decision

negative externality the costs imposed on a third party of a decision

positive externality the benefits to a third party of a decision

The Social Costs and Social Benefits of Decision Making

We have seen how the operation of markets is based on millions of decisions being made by individuals and groups. The invisible hand means that individuals and groups make decisions which are designed to maximize their individual or group welfare. In making these decisions there will be private costs and private benefits. In making a car journey, for example, a person incurs various private costs such as the fuel used in the journey, the wear and tear (depreciation) on the car, the contribution of any vehicle tax and insurance costs related to the journey that the individual has to pay.

In using their car the individual also gains a number of private benefits: convenience, warmth, the pleasure of driving, listening to music, not to mention getting to a destination relatively quickly. However, in deciding to make the journey the individual may not take into consideration the cost (or benefit) **to society** that is imposed as a result of that decision. An extra car on the road contributes to congestion, road wear and tear, emissions that the car gives off, the noise pollution and the increased risk of accident which may cause injury or even death to a third party. There may also be some social benefits of the decision; using a car means that there is an extra seat available for someone else to use on public transport, for example.

These social costs and benefits are not taken into consideration by the individual as they get into their car. There are costs which have to be borne by a third party. The cost of repairing damaged roads, the cost of dealing with accident and injury, delays caused as a result of congestion, the effects and costs of dealing with pollution and so on, all have to be borne by others – often the taxpayer. Equally, any social benefits arising from the decision are gained by those not party to the initial decision without them having to pay for the benefit derived.

In the presence of externalities, society's interest in a market outcome extends beyond the well-being of buyers and sellers who participate in the market; it also includes the well-being of bystanders who are affected indirectly. Because buyers and sellers neglect the external effects of their actions when deciding how much to demand or supply, the market equilibrium is not efficient when there are externalities. That is, the equilibrium fails to maximize the total benefit to society as a whole.

The release of dioxin into the environment, for instance, is a negative externality. Self-interested paper firms will not consider the full cost of the pollution they create and, therefore, will emit too much pollution unless the government prevents or discourages them from doing so.

Types of Externalities Externalities come in many varieties, as do the policy responses that try to deal with the market failure. Here are some examples:

- The exhaust from cars is a negative externality because it creates smog that other people have to breathe. Drivers do not take into consideration this externality and so tend to drive too much, thus increasing pollution. The government attempts to solve this problem by setting emission standards for cars. It may also tax petrol and vehicle ownership in order to reduce the amount that people drive.
- Restored historic buildings convey a positive externality because people who walk or drive by them can enjoy their beauty and the sense of history that these buildings provide. Building owners do not get the full benefit of restoration and, therefore, tend to discard older buildings too quickly. Many national governments respond to this problem by regulating the destruction of historic buildings and by providing tax incentives to owners who restore them.
- Barking dogs create a negative externality because neighbours are disturbed by the noise. Dog owners do not bear the full cost of the noise and, therefore, tend to take too few precautions to prevent their dogs from barking. The government may address this problem by making it illegal to 'disturb the peace'.
- Research into new technologies provides a positive externality because it creates knowledge that other people can use. Because inventors cannot capture the full benefits of their inventions, they tend to devote too few resources to research. The government addresses this problem partially through the patent system, which gives inventors an exclusive use over their inventions for a period of time.
- A programme of vaccination against a flu virus or any other communicative disease protects those who receive it from the risk of contracting the virus. Those who are not vaccinated, however, may receive some benefit too because the prevalence of the virus is lower and so there is a reduced risk that they will contract the illness. Health services also benefit because they do not have to devote resources to treating those with illnesses. Governments encourage vaccinations because there are positive benefits to society as a whole.

In each of these cases, some decision makers fails to take account of the external effects of their behaviour. The government responds by trying to influence this behaviour to protect the interests of bystanders.

EXTERNALITIES AND MARKET INEFFICIENCY

In this section we use the concepts of consumer and producer surplus and deadweight loss to examine how externalities affect economic well-being. The analysis shows precisely why externalities cause markets to allocate resources inefficiently.

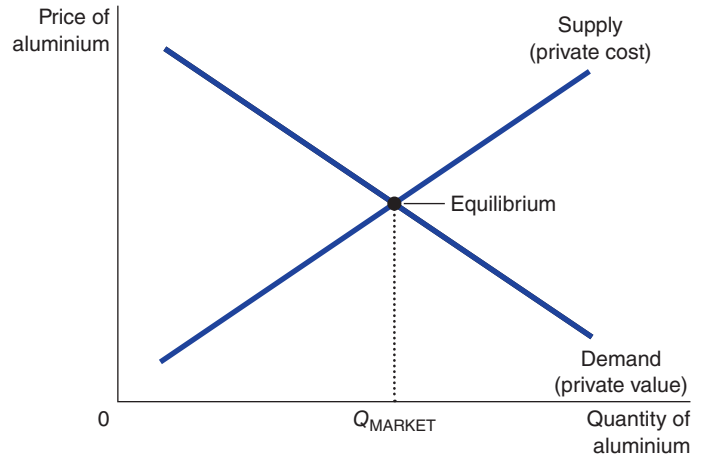
Welfare Economics: A Recap

To make our analysis concrete, we will consider a specific market – the market for aluminium. Figure 11.1 shows the supply and demand curves in the market for aluminium.

FIGURE 11.1

The Market for Aluminium

The demand curve reflects the value to buyers, and the supply curve reflects the costs of sellers. The equilibrium quantity, Q_{MARKET} , maximizes the total value to buyers minus the total costs of sellers. In the absence of externalities, therefore, the market equilibrium is efficient.



The supply and demand curves contain important information about costs and benefits. The demand curve for aluminium reflects the value of the benefits of aluminium to consumers, as measured by the prices they are willing to pay. At any given quantity, the height of the demand curve shows the willingness to pay of the marginal buyer. In other words, it shows the value to the consumer of the last unit of aluminium bought. Similarly, the supply curve reflects the costs of producing aluminium. At any given quantity, the height of the supply curve shows the cost of the marginal seller. In other words, it shows the cost to the producer of the last unit of aluminium sold. The demand and supply curves, therefore, reflect the private benefit to consumers and the private cost to suppliers.

In the absence of government intervention, the price adjusts to balance the supply and demand for aluminium. The quantity produced and consumed in the market equilibrium, shown as Q_{MARKET} in Figure 11.1, is efficient in the sense that it maximizes the sum of producer and consumer surplus. That is, the market allocates resources in a way that maximizes the total value to the consumers who buy and use aluminium minus the total costs to the producers who make and sell aluminium. At the equilibrium price, the value placed on the last unit of aluminium consumed by buyers is the same as the cost incurred by sellers of supplying that last unit.

Negative Externalities

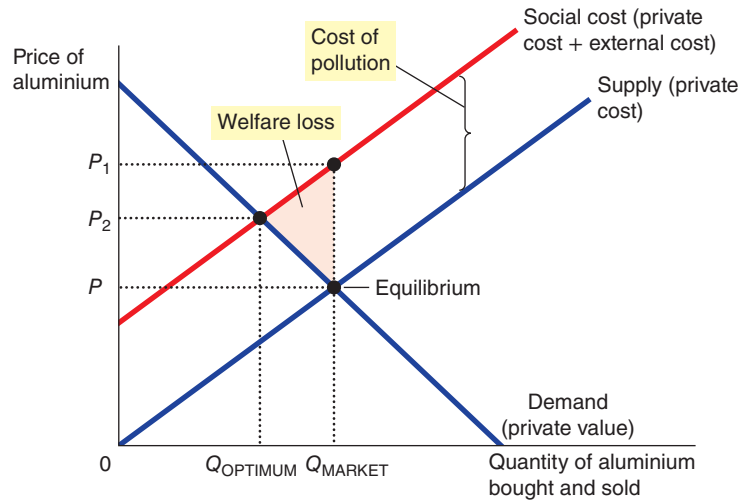
Now let's suppose that aluminium factories emit pollution: for each unit of aluminium produced, a certain amount of a pollutant enters the atmosphere. This pollutant may pose a health risk for those who breathe the air, it is a negative externality. There is a cost involved in dealing with the effects of the pollutant which may be the health care that those affected have to receive. This cost is not taken into consideration by producers of aluminium who only consider the private costs of production. How does this externality affect the efficiency of the market outcome?

Because of the externality, the cost to *society* of producing aluminium is larger than the cost to the aluminium producers. For each unit of aluminium produced, the *social (or external) cost* includes the private costs of the aluminium producers plus the costs to those bystanders affected adversely by the pollution. Figure 11.2 shows the social cost of producing aluminium. The social cost curve is above the supply curve because it takes into account the external costs imposed on society by aluminium producers. At every price the social cost is higher than the private cost, measured by the vertical distance between the social

cost curve and the private cost curve, so we can say that the social cost curve is the sum of the private costs and the social or external cost. The difference between these two curves reflects the social or external cost of the pollution emitted.

FIGURE 11.2**Pollution and the Social Optimum**

In the presence of a negative externality, such as pollution, the social cost of the good exceeds the private cost. The optimal quantity or socially efficient outcome, Q_{OPTIMUM} , is therefore smaller than the equilibrium quantity, Q_{MARKET} .



The Social Optimum or Socially Efficient Outcome At the market outcome Q_{MARKET} , consumers value the benefits of consuming this quantity of aluminium at OP . The true cost of Q_{MARKET} , however, is higher at OP_1 – the marginal consumer values aluminium at less than the social cost of producing it. The vertical distance OP_1P represents the welfare loss of producing Q_{MARKET} which is equal to the social cost of producing that output. The optimal amount of aluminium produced would be where the demand curve intersects the social cost curve at Q_{OPTIMUM} at a price of OP_2 . This is also referred to as the *socially efficient outcome*. This is the optimal amount of aluminium from the standpoint of society as a whole. At OP_2 consumers value the benefits of consuming aluminium at the same level as the private cost to suppliers and the cost to society as a whole. The socially efficient quantity is, therefore, lower than the private market outcome and the socially efficient price is higher reflecting the true value to society of the socially efficient market outcome.

The equilibrium quantity of aluminium, Q_{MARKET} , is larger than the socially optimal quantity, Q_{OPTIMUM} . The reason for this inefficiency is that the market equilibrium reflects only the private costs of production. In the market equilibrium, thus, reducing aluminium production and consumption below the market equilibrium level raises total economic well-being. We can measure changes in well-being by the welfare loss associated with different market outcomes. We can measure this welfare loss through summing the distance between the value placed on the consumption of aluminium and the social cost of production between the market outcome, Q_{MARKET} and the optimum outcome Q_{OPTIMUM} . The total amount is shown by the shaded triangle in Figure 11.2. This triangle is referred to as a *welfare triangle*.

If the market outcome is deemed to be inefficient from the standpoint of society as a whole, what can governments do to achieve a socially desirable outcome? The answer is to somehow force the decision maker to take into consideration some or all of the social costs of the decision. In our example, one way to do this would be to tax aluminium producers for each tonne of aluminium sold. The tax would shift the supply curve for aluminium upward by the size of the tax. If the tax accurately reflected the social cost of the pollution released into the atmosphere, the new supply curve would coincide with the social cost curve. In the new market equilibrium, aluminium producers would produce the socially optimal quantity of aluminium.

The use of such a tax is called **internalizing an externality** because it gives buyers and sellers in the market an incentive to take account of the external effects of their actions. By having to include the tax

into their decision making, the government is able to intervene to make the price signal more accurate. Aluminium producers would, in essence, take the costs of pollution into account when deciding how much aluminium to supply because the tax would make them pay for these external costs. The policy is based on one of the *Ten Principles of Economics*: people respond to incentives. Later in this chapter we consider other ways in which policymakers can deal with externalities.

internalizing an externality altering incentives so that people take account of the external effects of their actions

Positive Externalities

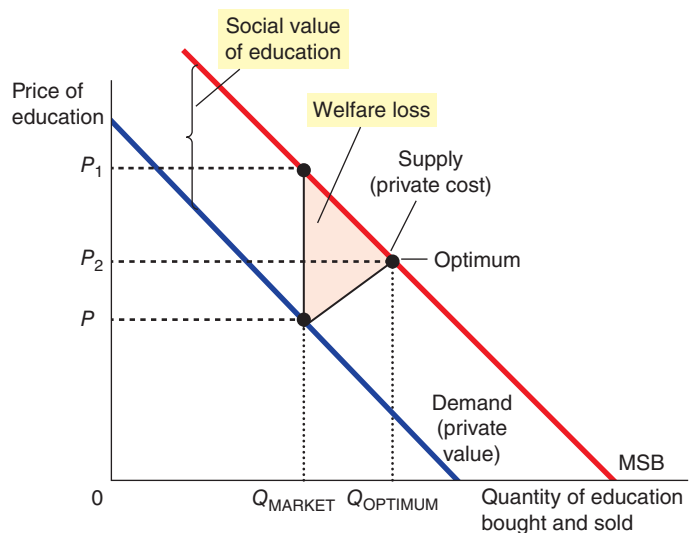
Although some activities impose costs on third parties, others yield benefits. For example, consider education. Education yields positive externalities because a more educated population leads to improved productivity and increases the potential for economic growth, which benefits everyone. Notice that the productivity benefit of education is not necessarily an externality: the consumer of education reaps most of the benefit in the form of higher wages and greater job mobility. But if some of the productivity benefits of education spill over and benefit other people, as is the case if economic growth is stimulated, then this effect would count as a positive externality as well.

The analysis of positive externalities is similar to the analysis of negative externalities. As Figure 11.3 shows, the demand curve does not reflect the value to society of the good. The value placed on an activity such as education is valued less by consumers than the total value to society. At the equilibrium market allocation of Q_{MARKET} the value of the private benefits to individuals of education is OP but the value to society as a whole is OP_1 . The vertical distance between OP and OP_1 is the value of the social benefits to society. Because the social value (or external benefit) is greater than the private value, the social value curve, or marginal social benefit (MSB) curve, lies above the demand curve. The MSB is the private value plus the external benefit to society at each price. At every price the benefit to society is greater than the private benefit, hence the social value curve lies to the right of the private benefit curve. The optimal quantity is found where the social value curve and the supply curve (which represents costs) intersect at a price of OP_2 . Hence, the socially optimal quantity is greater than the quantity determined by the private market and the price is higher than the private equilibrium price which reflects that the value of education is underpriced at market equilibrium. The welfare loss associated with the private market outcome at Q_{MARKET} is shown by the shaded triangle.

FIGURE 11.3

Education and the Social Optimum

In the presence of a positive externality, the social value of the good exceeds the private value. The optimal quantity, Q_{OPTIMUM} , is therefore larger than the equilibrium quantity, Q_{MARKET} .



Once again, the government can correct the market failure by inducing market participants to internalize the externality. The appropriate response in the case of positive externalities is exactly the opposite to the case of negative externalities. To move the market equilibrium closer to the social optimum, a positive externality requires a subsidy. In fact, that is exactly the policy many governments follow by heavily subsidizing education.

To summarize: negative externalities lead markets to produce a larger quantity than is socially desirable. Positive externalities lead markets to produce a smaller quantity than is socially desirable. To remedy the problem, the government can internalize the externality by taxing goods that have negative externalities and subsidizing goods that have positive externalities.

SELF TEST Give an example of a negative externality and a positive externality. ● Explain why market outcomes are inefficient in the presence of externalities.

PRIVATE SOLUTIONS TO EXTERNALITIES

We have discussed why externalities lead markets to allocate resources inefficiently and now we turn to how this inefficiency can be remedied. In practice, both private actors and public policymakers respond to externalities in various ways. All of the remedies share the goal of moving the allocation of resources closer to the social optimum. In this section we examine private solutions.

The Types of Private Solution

Although externalities tend to cause markets to be inefficient, government action is not always needed to solve the problem. In some circumstances, people can develop private solutions.

Social Norms of Moral Behaviour Sometimes the problem of externalities is solved with moral codes and social sanctions. Consider, for instance, why most people do not litter. Although there are laws against littering, these laws are not vigorously enforced. Most people do not litter just because it is the wrong thing to do. Advertising and parental guidance help us to distinguish what society accepts as a norm for behaviour. This moral injunction tells us to take account of how our actions affect other people. In economic terms, it tells us to internalize externalities.

Charities Many charities are established to deal with externalities. For example, Greenpeace, whose goal is to protect the environment, is a non-profit organization funded with private donations. As another example, colleges and universities sometimes receive gifts from alumni, corporations and foundations in part because education has positive externalities for society.

Self-interest The private market can often solve the problem of externalities by relying on the self-interest of the relevant parties. Sometimes the solution takes the form of integrating different types of business. For example, consider an apple grower and a beekeeper who are located next to each other. Each business confers a positive externality on the other: by pollinating the flowers on the trees, the bees help the orchard produce apples. At the same time, the bees use the nectar they get from the apple trees to produce honey. Nevertheless, when the apple grower is deciding how many trees to plant and the beekeeper is deciding how many bees to keep, they neglect the positive externality. As a result, the apple grower plants too few trees and the beekeeper keeps too few bees. These externalities could be internalized if the beekeeper bought the apple orchard or if the apple grower bought the beehives: both activities would then take place within the same firm, and this single firm could choose the optimal number of trees and bees. Internalizing externalities is one reason that some firms are involved in different types of business.

Social Contracts Another way for the private market to deal with external effects is for the interested parties to enter into a contract. In the foregoing example, a contract between the apple grower and the

beekeeper can solve the problem of too few trees and too few bees. The contract can specify the number of trees, the number of bees and perhaps a payment from one party to the other. By setting the right number of trees and bees, the contract can solve the inefficiency that normally arises from these externalities and make both parties better off. Joint ventures and partnering are good examples of where such contracts can generate positive externalities and improve efficiency.

The Coase Theorem

How effective is the private market in dealing with externalities? A famous result, called the **Coase theorem**, after British economist Ronald Coase, suggests that it can be very effective in some circumstances. According to the Coase theorem, if private parties can bargain *without cost* over the allocation of resources, then the private market will always solve the problem of externalities and allocate resources efficiently.

Coase theorem the proposition that if private parties can bargain without cost over the allocation of resources, they can solve the problem of externalities on their own

To see how the Coase theorem works, consider an example. Suppose that Sofie owns a dog named Brandy. Brandy barks and disturbs Lucas, Sofie's neighbour. Sofie gets a benefit from owning the dog, but the dog confers a negative externality on Lucas. Should Sofie be forced to send Brandy to the dogs' home or should Lucas have to suffer sleepless nights because of Brandy's barking?

Consider first what outcome is socially efficient. We might consider two options. One is to compare the benefit that Sofie gets from the dog to the cost that Lucas bears from the barking. If the benefit exceeds the cost, it is efficient for Sofie to keep the dog and for Lucas to live with the barking. The second option is that if the cost to Lucas exceeds the benefit to Sofie, then she should get rid of the dog. The problem is in valuing the respective costs and benefits.

According to the Coase theorem, the private market can reach the efficient outcome on its own by Lucas offering to pay Sofie to get rid of the dog. Sofie will accept the deal if the amount of money Lucas offers is greater than the benefit to her of keeping the dog.

By bargaining over the price, Sofie and Lucas can always reach the efficient outcome. For instance, suppose that Sofie gets a €500 benefit from the dog and Lucas bears an €800 cost from the barking. In this case, Lucas can offer Sofie €600 to get rid of the dog, and Sofie will gladly accept. Both parties are better off than they were before, and the efficient outcome is reached.

It is possible, of course, that Lucas would not be willing to offer any price that Sofie would accept. For instance, suppose that Sofie gets a €1,000 benefit from the dog and Lucas bears an €800 cost from the barking. In this case, Sofie would turn down any offer below €1,000, while Lucas would not offer any amount above €800. Therefore, Sofie ends up keeping the dog. Given these costs and benefits, however, this outcome is efficient.

So far, we have assumed that Sofie has the legal right to keep a barking dog. In other words, we have assumed that Sofie can keep Brandy unless Lucas pays her enough to induce her to give up the dog voluntarily. How different would the outcome be, on the other hand, if Lucas had the legal right to peace and quiet?

According to the Coase theorem, the initial distribution of rights does not matter for the market's ability to reach the efficient outcome. For instance, suppose that Lucas can legally compel Sofie to get rid of the dog. Although having this right works to Lucas's advantage, it probably will not change the outcome. In this case, Sofie can offer to pay Lucas to allow her to keep the dog. If the benefit of the dog to Sofie exceeds the cost of the barking to Lucas, then Sofie and Lucas will strike a bargain in which Sofie keeps the dog.

Although Sofie and Lucas can reach the efficient outcome regardless of how rights are initially distributed, the distribution of rights is not irrelevant: it determines the distribution of economic well-being. Whether Sofie has the right to a barking dog or Lucas the right to peace and quiet determines who pays whom in the final bargain. But, in either case, the two parties can bargain with each other and solve the externality problem. Sofie will end up keeping the dog only if the benefit exceeds the cost.

To sum up: the Coase theorem says that private economic actors can solve the problem of externalities among themselves. Whatever the initial distribution of rights, the interested parties can always reach a bargain in which everyone is better off and the outcome is efficient.

Why Private Solutions Do Not Always Work

Despite the appealing logic of the Coase theorem, private actors on their own often fail to resolve the problems caused by externalities. The Coase theorem applies only when the interested parties have no trouble reaching and enforcing an agreement. In the world, however, bargaining does not always work, even when a mutually beneficial agreement is possible.

Transaction Costs Sometimes the interested parties fail to solve an externality problem because of **transaction costs**, the costs that parties incur in the process of agreeing to and following through on a bargain. In our example, imagine that Sofie and Lucas speak different languages so that, to reach an agreement, they will need to hire a translator. If the benefit of solving the barking problem is less than the cost of the translator, Sofie and Lucas might choose to leave the problem unsolved. In more realistic examples, the transaction costs are the expenses not of translators but of the lawyers required to draft and enforce contracts.

transaction costs the costs that parties incur in the process of agreeing and following through on a bargain

Bargaining Problems At other times bargaining simply breaks down. The recurrence of wars and labour strikes shows that reaching agreement can be difficult and that failing to reach agreement can be costly. The problem is often that each party tries to hold out for a better deal. For example, suppose that Sofie gets a €500 benefit from the dog, and Lucas bears an €800 cost from the barking. Although it is efficient for Lucas to pay Sofie to get rid of the dog, there are many prices that could lead to this outcome. Sofie might demand €750, and Lucas might offer only €550. As they haggle over the price, the inefficient outcome with the barking dog persists.

Coordinating Interested Parties Reaching an efficient bargain is especially difficult when the number of interested parties is large because coordinating everyone is costly. For example, consider a factory that pollutes the water of a nearby lake. The pollution confers a negative externality on the local fishermen. According to the Coase theorem, if the pollution is inefficient, then the factory and the fishermen could reach a bargain in which the fishermen pay the factory not to pollute. If there are many fishermen, however, trying to coordinate them all to bargain with the factory may be almost impossible.

Asymmetric Information and the Assumption of Rational Behaviour There are two other key reasons why reaching an efficient bargain may not arise: asymmetric information and the assumption of rational behaviour. An example of the former is that Sofie and Lucas may not have perfect knowledge of the costs and benefits to each other of the barking dog. In such situations it becomes very difficult to negotiate an efficient outcome. Both parties have imperfect information about the situation of the other and so incentives may be distorted. Lucas, for example, might exaggerate the cost to him of the barking dog whilst Sofie does the same with regard to the benefits she gets from keeping her dog. The situation is further complicated by the existence of *free riders*. Lucas may not be the only person in the neighbourhood suffering from the barking dog but others may not live directly next to Sofie. These other 'victims' can benefit from any agreement that Sofie and Lucas arrive at but do not pay any of the costs of solving the problem. If Lucas is aware of this then why should he pay the full amount to solve the problem when others will also benefit but not contribute? If all victims think the same way then the problem will remain unsolved and there will be an inefficient outcome.

As regards the assumption of rational behaviour, we assumed that an efficient outcome could be found if Lucas offered €600 for Sofie to get rid of the dog. If Sofie were able to put a price on the value of the dog to her and this was €500, then it would be irrational for her not to accept the money to get rid of the dog.

The money could be used to secure something which gave greater value to her than the ownership of the dog. Of course, in real life such rational behaviour may be clouded by all sorts of behavioural and psychological influences that Sofie may not be able to value: the guilt she may feel in getting rid of the dog, the reactions of her friends and family, the sentimental value of the dog to her and so on.

When private bargaining does not work, the government can sometimes play a role. The government is an institution designed for collective action. In the polluting factory example above, the government can act on behalf of the fishermen, even when it is impractical for the fishermen to act for themselves. In the next section, we examine how the government can try to remedy the problem of externalities.

SELF TEST Give an example of a private solution to an externality. ● What is the Coase theorem? ● Why are private economic actors sometimes unable to solve the problems caused by an externality?

Positional Externalities

In many walks of life people and businesses strive to improve themselves in some way. When decisions are made which are intended to improve one person's performance (or payoff) relative to another, there are effects on the other people. For example, if a professional squash player wants to improve her performance, she might hire a sports psychologist to advise her on getting that extra edge that might make a difference to their performance and win percentage. As a result of the squash player making this decision, other squash players' performances (their payoff) are affected negatively. When a payoff of one individual or business is dependent on their performance relative to a rival this is called a **positional externality**. The improvement in performance of the squash player who hires the sports psychologist in terms of the payoff she receives (i.e. winning) will of necessity worsen her rivals' payoffs (i.e. they will now not win). What is the incentive for all other professional squash players? It is to also hire a sports psychologist and see if they can make up that competitive gap and possibly improve it. If all squash players do this, however, the overall effect is likely to be neutral and so the overall benefit of professional squash players as a collective is zero despite the additional costs incurred. Such an outcome, therefore is inefficient.

positional externality a situation which exists when the payoff to one individual is dependent on their relative performance to others

We can see the existence of positional externalities in the world of education and business. To get to university, individuals require certain grades in examinations taken at the end of secondary or high school/college education. Universities demand ever higher grades in an effort to select the highest performing students and those with the most academic potential. As a result, individual schools and colleges might try to find ways to help their students improve their grades relative to other schools and colleges.

However, the result is that every school and college invests scarce resources into extra revision classes, sending teachers onto courses to better understand the examination systems and other measures to try and gain that competitive advantage for their students. What universities are faced with is ever increasing numbers of students applying with similar qualifications thus making differentiation more difficult.

Firms looking to recruit the best workers might request particular skills and qualifications. This may be the requirement for applicants to have a Master's degree or to have had internships at relevant businesses or to have developed people and communication skills. Individual applicants have an incentive to invest time and resources in gaining these qualifications and skills to try and put themselves ahead of rival applicants. The overall effect is that all applicants present themselves with Master's degrees and any amount of experience and skills so the outcome is limited benefit to the recruiting firm and considerable cost for the applicants.

The existence of positional externalities which leads to individuals investing in a series of measures designed to gain them an advantage but which simply offset each other is referred to as **positional arms races**.

positional arms races a situation where individuals invest in a series of measures designed to gain them an advantage but which simply offset each other

PUBLIC POLICIES TOWARDS EXTERNALITIES

When an externality causes a market to reach an inefficient allocation of resources, the government can respond in one of two ways. *Command-and-control policies* regulate behaviour directly. *Market-based policies* provide incentives so that private decision makers will choose to solve the problem on their own through manipulation of the price signal.

Regulation

The government can remedy an externality by making certain behaviours either required or forbidden. For example, it is a crime in any European country to dump poisonous chemicals into the water supply. In this case, the external costs to society far exceed the benefits to the polluter. The government therefore institutes a command-and-control policy that prohibits this act altogether.

In most cases of pollution, however, the situation is not this simple. Despite the stated goals of some environmentalists, it would be impossible to prohibit all polluting activity. For example, virtually all forms of transport – even the horse – produce some undesirable polluting by-products. But it would not be sensible for the government to ban all transport. Thus, instead of trying to eradicate pollution altogether, society has to weigh the costs and benefits to decide the kinds and quantities of pollution it will allow.

Environmental regulations can take many forms. Sometimes the government may dictate a maximum level of pollution that a factory may emit. Other times the government requires that firms adopt a particular technology to reduce emissions. In all cases, to design good rules, the government regulators need to know the details about specific industries and about the alternative technologies that those industries could adopt. This information is often difficult for government regulators to obtain.

Pigovian Taxes and Subsidies

Instead of regulating behaviour in response to an externality, the government can use market-based policies to align private incentives with social efficiency. For instance, as we saw earlier, the government can internalize the externality by taxing activities that have negative externalities and subsidizing activities that have positive externalities. Taxes enacted to correct the effects of negative externalities are called **Pigovian taxes**, after the English economist Arthur Pigou (1877–1959), an early advocate of their use.

Pigovian tax a tax enacted to correct the effects of a negative externality

Economists usually prefer Pigovian taxes over regulations as a way to deal with pollution because such taxes can reduce pollution at a lower cost to society. To see why, let us consider an example.

Suppose that two factories – a paper mill and a steel mill – are each dumping 500 tonnes of effluent into a river each year. The government decides that it wants to reduce the amount of pollution. It considers two solutions:

- *Regulation.* The government could tell each factory to reduce its pollution to 300 tonnes of effluent per year.
- *Pigovian tax.* The government could levy a tax on each factory of €50,000 for each tonne of effluent it emits.

The regulation would dictate a level of pollution, whereas the tax would give factory owners an economic incentive to reduce pollution. The intention of the tax would be to encourage firms to reduce pollution up to

the point where the marginal abatement cost is equal to the tax rate imposed. The **marginal abatement cost** is the cost expressed in terms of the last unit of pollution not emitted (abated). Which solution do you think is more likely to bring about the desired objective of reducing pollution?

marginal abatement cost the cost expressed in terms of the last unit of pollution not emitted (abated)

Most economists would prefer the tax. They would first point out that a tax is just as effective as a regulation in reducing the overall level of pollution. The government can achieve whatever level of pollution it wants by setting the tax at the appropriate level. The higher the tax, the larger the reduction in pollution. Indeed, if the tax is high enough, the factories will close down altogether, reducing pollution to zero.

The reason why economists would prefer the tax is that it reduces pollution more efficiently. The regulation requires each factory to reduce pollution by the same amount, but an equal reduction is not necessarily the least expensive way to clean up the water. It is possible that the paper mill can reduce pollution at lower cost than the steel mill. If so, the paper mill would respond to the tax by reducing pollution substantially to avoid the tax, whereas the steel mill would respond by reducing pollution less and paying the tax.

In essence, the Pigovian tax places a price on the right to pollute. Just as markets allocate goods to those buyers who value them most highly, a Pigovian tax allocates pollution to those factories that face the highest cost of reducing it. Whatever the level of pollution the government chooses, it can achieve this goal at the lowest total cost using a tax.

Economists also argue that Pigovian taxes are better for the environment. Under the command-and-control policy of regulation, the factories have no reason to reduce emission further once they have reached the target of 300 tonnes of effluent. By contrast, the tax gives the factories an incentive to develop cleaner technologies, because a cleaner technology would reduce the amount of tax the factory has to pay.

Pigovian taxes are unlike most other taxes. Most taxes distort incentives and move the allocation of resources away from the social optimum. The reduction in economic well-being – that is, in consumer and producer surplus – exceeds the amount of revenue the government raises, resulting in a deadweight loss. By contrast, when externalities are present, society also cares about the well-being of the bystanders who are affected. Pigovian taxes correct incentives for the presence of externalities and thereby move the allocation of resources closer to the social optimum. Thus, while Pigovian taxes raise revenue for the government, they also enhance economic efficiency.

Despite the logic of Pigovian taxes, examples of pollution taxes are scarce. Some economists point out that what constitutes a pollution tax depends on how environmental tax systems are defined. There are, however, some problems associated with such taxes, not least identifying the appropriate rate to levy. In addition there are political problems associated with levying Pigovian taxes. The cost of levying and administering these taxes might be higher compared to regulation.

Tradable Pollution Permits

Returning to our example of the paper mill and the steel mill, let us suppose that the government decides to adopt regulation and requires each factory to reduce its pollution to 300 tonnes of effluent per year. Then one day, after the regulation is in place and both mills have complied, the two firms go to the government with a proposal. The steel mill wants to increase its emission of effluent by 100 tonnes. The paper mill has agreed to reduce its emission by the same amount if the steel mill pays it €5 million. Should the government allow the two factories to make this deal?

From the standpoint of economic efficiency, allowing the deal is good policy. The deal must make the owners of the two factories better off, because they are voluntarily agreeing to it. Moreover, the deal does not have any external effects because the total amount of pollution remains the same. Thus, social welfare is enhanced by allowing the paper mill to sell its right to pollute to the steel mill.

The same logic applies to any voluntary transfer of the right to pollute from one firm to another. If the government allows firms to make these deals, it will, in essence, have created a new scarce resource: pollution permits. A market to trade these permits will eventually develop, and that market will be governed

by the forces of supply and demand. The invisible hand will ensure that this new market efficiently allocates the right to pollute. The firms that can reduce pollution only at high cost will be willing to pay the most for the pollution permits. The firms that can reduce pollution at low cost will prefer to sell whatever permits they have.

One advantage of allowing a market for pollution permits is that the initial allocation of pollution permits among firms does not matter from the standpoint of economic efficiency. The logic behind this conclusion is similar to that behind the Coase theorem. Those firms that can reduce pollution most easily would be willing to sell whatever permits they get, and those firms that can reduce pollution only at high cost would be willing to buy whatever permits they need. As long as there is a free market for the pollution rights, the final allocation will be efficient whatever the initial allocation.

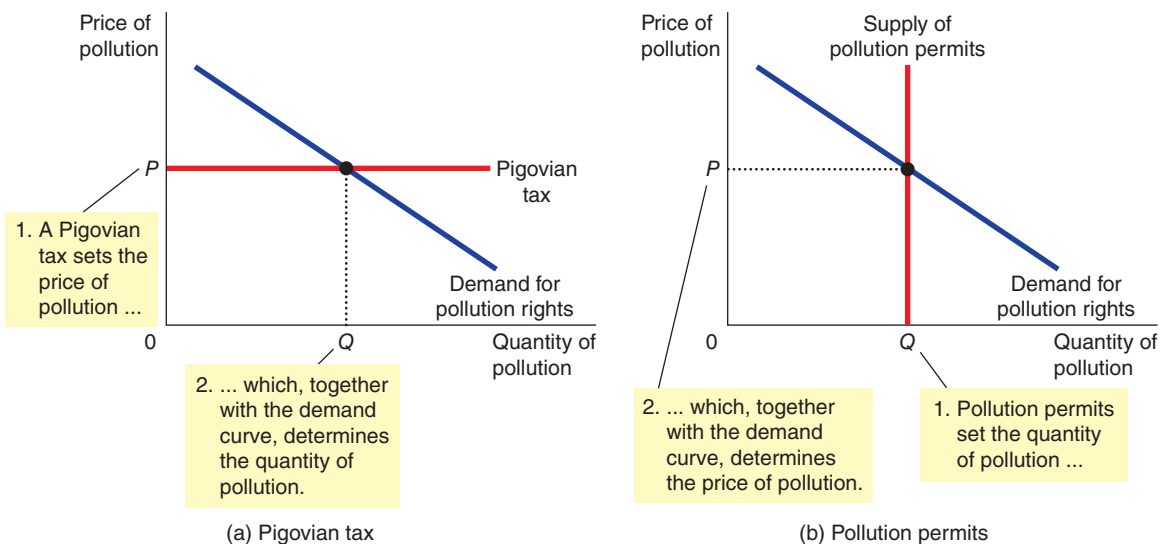
Although reducing pollution using pollution permits may seem quite different from using Pigovian taxes, in fact the two policies have much in common. In both cases, firms pay for their pollution. With Pigovian taxes, polluting firms must pay a tax to the government. With pollution permits, polluting firms must pay to buy the permit. (Even firms that already own permits must pay to pollute: the opportunity cost of polluting is what they could have received by selling their permits on the open market.) Both Pigovian taxes and pollution permits internalize the externality of pollution by making it costly for firms to pollute.

The similarity of the two policies can be seen by considering the market for pollution. Both panels in Figure 11.4 show the demand curve for the right to pollute. This curve shows that the lower the price of polluting, the more firms will choose to pollute. In panel (a) the government uses a Pigovian tax to set a price for pollution. In this case, the supply curve for pollution rights is perfectly elastic (because firms can pollute as much as they want by paying the tax), and the position of the demand curve determines the quantity of pollution. In panel (b) the government sets a quantity of pollution by issuing pollution permits. The level at which this quantity is set is crucial. In this case, the supply curve for pollution rights is perfectly inelastic (because the quantity of pollution is fixed by the number of permits), and the position of the demand curve determines the price of pollution. Hence, for any given demand curve for pollution, the government can achieve any point on the demand curve either by setting a price with a Pigovian tax or by setting a quantity with pollution permits.

FIGURE 11.4

The Equivalence of Pigovian Taxes and Pollution Permits

In panel (a) the government sets a price on pollution by levying a Pigovian tax, and the demand curve determines the quantity of pollution. In panel (b) the government limits the quantity of pollution by limiting the number of pollution permits, and the demand curve determines the price of pollution. The price and quantity of pollution are the same in both cases.



In some circumstances, however, selling pollution permits may be better than levying a Pigovian tax. Suppose the government wants no more than 600 tonnes of effluent to be dumped into the river. But, because the government does not know the demand curve for pollution, it is not sure what size tax would achieve that goal. In this case, it can simply auction off 600 pollution permits. The auction price would yield the appropriate size of the Pigovian tax.

The idea of the government auctioning off the right to pollute may at first sound like a creature of some economist's imagination. And, in fact, that is how the idea began. But a number of governments around the world have used such a system as a way to control pollution. In 2002, European Union environment ministers unanimously agreed to set up a market to trade pollution permits for carbon dioxide (CO₂), the main so-called greenhouse gas of concern. Pollution permits, like Pigovian taxes, are increasingly being viewed as a cost-effective way to keep the environment clean.

PUBLIC/PRIVATE POLICIES TOWARDS EXTERNALITIES

Property Rights

In some cases, private solutions to externalities can occur but need some form of legal back-up to be able to work. One such example is the establishment of **property rights**. Property rights refer to the exclusive right of an individual, group or organization to determine how a resource is used. The existence of well-established property rights enshrined in law allows the owners of that property to be able to use it as they see fit and to also have some protection in law if their rights are infringed in any way.

property rights the exclusive right of an individual, group or organization to determine how a resource is used

To see how this works, let us take a simple example. Lothar is the legally recognized owner of a Mercedes Benz car. Max is on his way home from a night out with his friends and as a prank lets down the tyres on Lothar's car. Close circuit TV cameras capture Max in the act and he is arrested for criminal damage. Lothar has the right to prosecute Max for the damage caused to his property and can expect to receive some money from Max to pay for the damages caused. The damage might not only be valued in terms of the cost of re-inflating the tyres but also for the estimated cost (decided by the courts) to Lothar of him having to miss a meeting early the next morning because he could not use his car.

It could be argued that in some cases, the market fails to allocate resources efficiently because property rights are not well established. That is, some item of value does not have an owner with the legal authority to control it. For example, although no one doubts that the 'good' of clean air or national defence is valuable, no one has the right to attach a price to it and profit from its use. A factory pollutes the air too much because no one owns the air which is polluted and can charge the factory for the pollution it emits. The market does not provide for national defence because no one can charge those who are defended for the benefit they receive.

Government Solutions to the Absence of Property Rights When the absence of property rights causes a market failure, the government can potentially solve the problem. Sometimes, as in the sale of pollution permits, the solution is for the government to help define property rights and thereby unleash market forces. In some countries, common resources such as rivers, are put under the ownership of an agency established by the government. The agency is given the right to ownership of the rivers and so can take action against those that cause damage to the river in some way. Other times, as in the restriction on hunting seasons, the solution is for the government to regulate private behaviour. Still other times, as in the provision of national defence, the solution is for the government to supply a good that the market fails to supply. In all cases, if the policy is well planned and well run, it can make the allocation of resources more efficient and thus raise economic well-being.

In order for any economy to work efficiently, a system of property rights has to be established and understood. This is not as easy as it sounds, however.

With things such as rivers, streams, land and air it is less easy to establish who are the legal owners. If some system could be devised whereby the ownership of property could be established then those that cause damage to that property can be brought to book! Extending property rights, therefore, might be one area where externalities can be internalized. For example, if property rights over the air that we breathe can be extended, then any firm polluting that air (in whatever way, noise, smell, smoke, etc.) could face prosecution for doing so. The threat of prosecution is sufficient to act as an incentive to find ways of not polluting the air. This might mean that a notional property zone is established above and around privately owned properties where the owner of the property also 'owns' the air above and around it. If that air is then polluted in some way, the owner can seek legal redress.

Extension of property rights also means that the owner of the property (which can be intellectual as well as physical) can also exercise the right to sell or share that property if they so wish at some mutually agreeable price. Extending property rights allows individuals, groups and organizations to be able to arrive at efficient solutions. If, for example, an individual was assigned property rights for the air one kilometre above their property, then if a nearby factory wanted to pollute that air they would have to enter into negotiations with the house owner to do so at some mutually agreeable price. The resulting right to pollute could also be sold to another party. A more developed system of property rights can, therefore, improve well-being and it has been identified as playing a crucial role in good governance, particularly relevant for developing countries to be able to attract the sort of inward investment that will help their economies to grow.

Difficulties in Establishing Property Rights There are problems with extending property rights, however. How do we apportion rights to such things as air, the seas, rivers and land? The cost of establishing property rights and getting international agreement on what they entail is considerable and may counteract the social benefits they might provide. If property rights were extended to the volume of air 1 km above a person's property, imagine the complexity of the negotiations that would have to be carried out with any business nearby, or airlines and the military for the right to share that air! Property owners may also have insufficient knowledge about their rights and exactly what they mean; it is also not a costless exercise to prove that property rights have been violated.

In the music industry the complexities of property rights have been the subject of debate and countless lawsuits in recent years. It not only relates to the issues of file sharing, pirating, copying CDs for personal use and downloading, but also to the artists themselves and the rights to the music that they have written and performed. Intellectual property law is an incredibly complex area and different countries interpret property rights in different ways making any international agreement even more difficult.

Despite the complexities, there have been efforts to extend property rights to help bring social benefits. In many parts of Europe, property rights over public spaces such as national parks, rivers and seas have meant that environmental laws can be established and enforced and this has led to an improvement in well-being for millions who are able to use these spaces, enjoy cleaner rivers and exploit the resources of the sea.

Control of Positional Arms Races

To reduce the instances of positional arms races, an incentive must exist to prevent the investment in attempts to gain some benefit which is ultimately mutually offsetting. This could take place through legislation to ban particular types of performance enhancing drug in sport, for example, or through some form of informal agreement between participants or through the establishment of a social norm on behaviour which becomes accepted by most participants.

In other cases some form of legally binding arbitration agreement might be made where participants agree to an external body overseeing a dispute and where all parties agree to abide by the decision made by this body.

CASE STUDY Carbon Trading Permits

The European Union has prided itself on being a leader in setting up carbon trading through the European Trading Scheme (ETS). The EU sets limits on the amount of carbon certain large industries like steel, cement and power generation are allowed to emit over a period of time. Permits are then issued to these industries and they have to work to meet their allowance. If they manage to reduce their carbon emissions below their permitted levels then they can sell the excess permits to others who may have had more difficulty meeting their targets.

Whilst such methods are, in theory, possible solutions to the problem of this type of pollution, the practice has seen different outcomes. The EU was accused of reducing the effectiveness of the system by *giving* permits to industry rather than auctioning them off. As a result there was a surplus of permits on the market and their price dropped. The effectiveness of the emissions trading scheme in reducing carbon emissions, and acting as an incentive to producers to find more socially efficient ways of operating, was called into question. The end of 2007 brought in the next phase of the scheme and it was hoped that the EU would learn the lessons from its mistakes.

The new limits, set to cover the period between 2008 and 2012, planned to tighten the amount of carbon pollution that can be emitted. In 2007 emissions of carbon were up by 1 per cent. However, the amount of carbon emitted was below targets set by the EU: 1.88 billion tonnes as against a target of 1.90 billion tonnes. It seems that France and Germany had been successful in having emissions below target levels whilst the UK, Spain and Italy were producing above their target levels. The overall figures, however, imply that tougher limits are required and that there would be a surplus of permits on the market, driving down prices.

The new targets set by the EU sought to reduce emissions by 9 per cent between 2008 and 2012. In addition, lessons from the first phase were learned and there were a larger number of permits that had to be bought by industry rather than being given to them. The new targets initially had some effect. The price of permits on the market began to rise. Since the beginning of 2008, when the new targets came into operation, permits were trading above €20 per tonne and analysts were expecting the price to rise further as the year progressed.

However, the recession in 2008 changed things. The global downturn led to a reduction in output and so firms were better able to meet their carbon emission limits. This also meant they were able to sell their permits and so the market saw a significant rise in the supply of permits which pushed down prices from a high of €30 to around €12 per permit. The Climate Change Summit, held in December 2009 in Copenhagen, was supposed to deliver a global agreement on carbon emissions and analysts were anticipating the market for carbon permits to rise significantly as a result of an announcement cutting carbon emissions. The failure to reach any binding agreement at the summit kept the market depressed and by early February 2010, prices were hovering at around €13 per permit.

At this level, the price of a permit does not present a sufficient incentive for firms to invest in technology to reduce carbon emissions. For firms to be prepared to incur the costs of investing in new technologies, the price of carbon permits must be high enough to give an incentive to divert resources to developing more efficient production methods and other technologies such as carbon capture and storage. If the cost of such investment is above the level a firm has to pay for a permit then there is no incentive to invest. Analysts have suggested that prices of permits need to be between €30 and €50 each to begin to have any effect on such investment; at these levels the opportunity cost of buying more permits to emit carbon starts to become too high. The market, it seems, has some way to go before it begins to deliver the social benefits hoped for.

How does a village become carbon neutral? What incentives have to be in place for the residents to subscribe to such an aim?



Objections to the Economic Analysis of Pollution

Some environmentalists argue that it is in some sense morally wrong to allow anyone to pollute the environment in return for paying a fee. Clean air and clean water, they argue, are fundamental human rights that should not be debased by considering them in economic terms. How can you put a price on clean air and clean water? The environment is so important, they claim, that we should protect it as much as possible, regardless of the cost.

Economists have little sympathy with this type of argument. To economists, good environmental policy begins by acknowledging the first of the *Ten Principles of Economics*: people face trade-offs. Certainly, clean air and clean water have value. But their value must be compared to their opportunity cost – that is, to what one must give up to obtain them. Eliminating all pollution is impossible. Trying to eliminate all pollution would reverse many of the technological advances that allow us to enjoy a high standard of living. Few people would be willing to accept poor nutrition, inadequate medical care or shoddy housing to make the environment as clean as possible.

Economists argue that some environmental activists hurt their own cause by not thinking in economic terms. A clean environment is a good like other goods. Like all normal goods, it has a positive income elasticity: rich countries can afford a cleaner environment than poor ones and, therefore, usually have more rigorous environmental protection. In addition, like most other goods, clean air and water obey the law of demand: the lower the price of environmental protection, the more the public will want. The economic approach of using pollution permits and Pigovian taxes reduces the cost of environmental protection and should, therefore, increase the public's demand for a clean environment.

SELF TEST A glue factory and a steel mill emit smoke containing a chemical that is harmful if inhaled in large amounts. Describe three ways the town government might respond to this externality. What are the pros and cons of each of your solutions?

GOVERNMENT FAILURE

In this chapter we have looked at how government can intervene to improve market outcomes. This of course is based on an assumption which is that government *can* improve market outcomes. This further implies that government decision making is based on high quality information and positive, rather than normative, analysis of problems and solutions. The reality is that government decision making is often flawed and not based on perfect information or rational, positive analysis. We have assumed that government intervention in the market is made to improve the efficiency and equity of market outcomes because the market left to itself somehow does not deliver those outcomes.

Governments, however, are made up of human beings who are invested with power to make decisions. What we may assume to be economic decisions being taken by government inevitably become political decisions. Politics is about power – who wields that power and how power is brought to bear on individuals and groups within government will affect decision making in ways which may not always amount to efficient nor equitable outcomes despite what politicians may have us believe. In considering these outcomes we look at the benefits to people of government decision making in relation to the costs. If the benefits are greater than the costs it can be argued that this can be deemed 'efficient'. However, there might be instances where the benefits accrue to a small number of people but the costs are spread across large sections of the population. In those circumstances it can be argued that the market outcome is inefficient. When governments make decisions that conflict with economic efficiency it is termed **government failure**. No government decision can be taken at face value without considering the politics behind it. In the final section of this chapter we will look at some aspects of government failure.

government failure a situation where political power and incentives distort decision making so that decisions are made which conflict with economic efficiency

Public Choice Theory

Governments are urged to step in to help improve market outcomes or solve perceived problems in many different areas: gun crime, drugs, prisons, poverty, housing, health, education, obesity, ultra-thin models, binge-drinking, racism, military intervention, terrorism, famine and so on. Ostensibly, government intervenes to act in the public interest and to try and improve market outcomes. The **public interest** can be defined as a principle based on making decisions to maximize the benefits gained from decisions to the largest number of people at minimum cost.

The circumstances under which governments intervene in markets might not be one based on rational analysis of the choices available but as a reaction to public pressure or moral panic spread by news organizations. Moral panic can lead to excessive pressure being put on governments and in such situations decisions can be made not on the basis of rationality and efficiency but on placating some individual or group self-interest. There are three key actors in public choice theory: voters, law-makers or politicians and bureaucrats.

public interest making decisions based on a principle where the maximum benefit is gained by the largest number of people at minimum cost

Decision making under these circumstances comes under the heading of public choice theory. **Public choice theory** is about the analysis of governmental behaviour, and the behaviour of individuals who interact with government.

public choice theory the analysis of governmental behaviour, and the behaviour of individuals who interact with government

The theory developed when economists looked at some apparent contradictions in human behaviour. Any decision will involve some sort of cost, so there will be some people that will be affected adversely by those decisions (the 'losers'). If those people are in the minority and the benefits to the majority (the 'winners') outweigh those costs, a decision might then be regarded as acting in the public interest. However, if the winners are in the minority and the costs are borne by the majority losers, then government failure might exist.

An Example: Road Congestion One solution for road congestion is to make people pay for the use of the roads. If this results in a reduction in road usage, or a more efficient use of the roads, there will not only be widespread benefits for road users, but also for the environment as a whole. However, if there is a vocal group that are very much against road pricing which have political power and influence (they might be backed by an influential newspaper, for example) this group might be able to use their political power to get the policy of road pricing abandoned. They may have decided to champion the cause of the road user, going so far as to provide car stickers showing support for the abandonment of road pricing. Would this be an efficient outcome?

The Invisible Hand versus Public Interest

Public choice theory developed out of an economics tradition that stems from Adam Smith's invisible hand. However, despite this seemingly clear support for self-interest, Smith did spend time discussing government in *The Wealth of Nations*. That discussion was based around the understanding of a moral concern with public interest.

The basis of public choice theory, therefore, centres on the behaviour of people as individuals in comparison to the behaviour of those individuals when they become political animals. Can an individual put aside their personal feelings and preferences and become transformed to understand and appreciate the broader public perspective when they are in government?

Public choice theory tries to look at the economic analysis of human behaviour as individuals and transfers this analysis to political science. What public choice theory looks at are cases where that individual interest leads to decisions and the allocation of resources which may not be the most efficient allocation.

Voter Incentives

Voters are asked to make choices in a democratic political system and to choose politicians to represent their views in government. It is assumed that voters will make their choices based on self-interest – the party or politician who offers the promise of the most benefits at the least cost to the voter. However, voters also know that their individual vote counts for little in the grand scheme of things (it is unlikely to make the difference between a government getting into power or not) and as a result they have little incentive to gather information on which to make an informed decision. This is called the **rational ignorance effect**.

rational ignorance effect the tendency of a voter to not seek out information to make an informed choice in elections

What information the voter does rely on to make their decision comes either from their parents (there is some evidence that voting follows such patterns), from TV coverage and from leaflets supplied by political parties. Such information is likely to be heavily biased and lacking in detail on the huge range of political decisions which any government has to make. As a result, the rational ignorance effect is reinforced. The effect also helps to explain why fewer people in some countries are taking the time to vote in elections; they simply do not see that their vote makes any difference and so there is little incentive for them to do so.

Politician Incentives

Politicians might claim that they have entered politics to fulfil a burning desire to act in the public interest. The more cynical might suggest that the actions and decisions made by politicians are motivated by attracting votes; after all, without votes politicians are not in a position to do anything even if they are genuinely motivated by the public interest.

It is in the interest of politicians to reflect the interests of the local communities they are seeking to serve because if they do they are more likely to attract votes and as a result get elected and re-elected. Some politicians can rely on the rational ignorance effect and the fact that the communities they represent have a large amount of voter inertia – in other words, voters are aligned to a political party so no matter what the politician does (within reason of course) it is highly unlikely they will not get re-elected.

Bureaucrat Incentives

Governments cannot survive without bureaucrats to administer government, provide advice and carry out the legislative programme. Civil servants, especially those in senior positions, are able to wield extensive power as a result and it is highly likely that they will seek to represent the interests of the particular agency or government department they happen to be working for. Those interests might involve protecting or securing larger departmental budgets but might also involve seeking career progression or recognition for the work they are doing. These interests are not necessarily aligned to economic efficiency; voters, for example, might want to see overseas aid or funding for the arts cut in times of domestic economic difficulty and the resources diverted to helping those in need in the domestic economy. Such a policy might have some economic merit but for the particular departments such a policy would be highly damaging to the politicians and bureaucrats involved and so there might exist a conflict of interest which might not lead to an efficient economic outcome.

The Special-Interest Effect

How might government failure manifest itself? Public choice theory likens politicians to a business. Imagine that a business produces a good which does not meet customer needs. It is very likely to fail. Politicians are like products: if they do not meet the customer's (voters) needs, they fail – they are not re-elected at the next election. The obvious behaviour, therefore, is to do what the consumer (the electorate) wants. What the electorate wants is not always clear. What is clear is that those who make the most noise and are the most organized are likely to be the ones that attract the most media attention or have access to the political decision makers whether they are politicians or bureaucrats and these may be the people who politicians listen to.

Individuals have specific knowledge about certain issues that are closely related to them. As a result of this information, special-interest groups that represent these views tend to develop and can exercise power through lobbying or cultivating close relationships. As a result, politicians who are able to influence decision making may be more inclined to listen to these groups and base their decisions on what these groups are saying. The **special-interest effect** may lead to minorities gaining significant benefits but the cost is borne by the population as a whole; the benefits to the winners are massively outweighed by the cost borne by the losers. If the value of the benefits to the special-interest group is less than the costs to the population as a whole, then this is an example of inefficiency.

special-interest effect where benefits to a minority special-interest group are outweighed by the costs imposed on the majority

In many cases it might be relatively easy for politicians to align themselves to a special-interest group, especially if that group does exercise political power by having connections with key media organizations or lobby groups. The incentives to align might not only be publicity but also the possibility of accessing funds to help support future election campaigns and as a result the incentive to support these organized groups is considerably stronger than the incentive to align with disorganized rationally ignorant voters.

Logrolling **Logrolling** is a term used to describe vote trading in government. Logrolling is an aspect of government failure that helps reinforce the special-interest effect. A voting member of the government (it could be the House of Commons in the UK, a member of the European Parliament or any other legislative chamber) will vote for something which they do not really support or believe in on the understanding that another member will vote in support on something that they do feel passionate about and want to support.

logrolling the agreement between politicians to exchange support on an issue

The argument for logrolling is that decisions or laws that affect relatively small groups of people to a significant extent can be secured, when, as a general rule, such laws would not be passed as it might not have a wide enough effect on society as a whole. The benefits to the group might still be more significant than the costs of implementing the policy or law. Without logrolling, those benefits would not be gained.

The problem with logrolling comes when the benefits to the winners are negligible in comparison to the costs to the losers. In a complex web of agreements and deals over voting, it is quite possible that the net gain to society is less than the costs imposed in achieving those net gains. In other words, resources are allocated inefficiently and the public interest is not maximized. One excellent example of where this might occur in reality is the whole issue of agricultural support – subsidies. The vast majority of the public know that subsidies exist but do not really understand the complexities of the issue or the effect it has on them.

For farmers, the issue is a very real one and there are plenty of lobbying groups that have particular self-interests – be it in dairy farming, arable farming, livestock and so on. The net effect of this web of agricultural support mechanisms in place throughout the world on society as a whole, represents a significant

misallocation of resources. The costs of agricultural support are less than the benefits. The benefits to those who receive them, however, are very important and might even mean the difference between being in business or not.

Rent Seeking

‘Rent’ in this context refers to the income some individual or group receives from an activity. The rents concerned do not always have positive social benefits; in fact, they are likely to have a negative social impact. **Rent seeking** refers to cases where resources are allocated to provide rents for individuals or groups and where those rents have negative social value.

rent seeking where individuals or groups take actions to redirect resources to generate income (rents) for themselves or the group

For example, in the US in 2003, President Bush announced tariffs would be imposed on steel imports to the US. The announcement was warmly welcomed by a steel industry that was struggling to compete with imports. The benefits of the tariffs to this group might be extensive, not least in the jobs that might be saved as a result. However, when balanced out against the wider effects, these benefits might pale into insignificance. There were threats of retaliation by other nations affected by the tariffs. This led to a reduction in demand for US-manufactured goods – along with potential job losses across a wide range of industries that might have used steel but also who had nothing to do with the steel industry. In addition, the tariffs meant that steel prices to US users rose – either through having to pay the tariff price for imported steel, or the switch to more expensive US producers. The ultimate effect was that the rents to the steel industry in the US could ultimately be argued to have had a negative social value.

What rent seeking amounts to is some group being able to influence policy to the extent that they are able to gain favours which take wealth from others and transfer that wealth to themselves. Whether this wealth transfer is economically efficient or not is not of concern to the group. Indeed it can also be argued that if rent seeking is successful these groups will divert more resources to such an activity rather than either solving the problems they are experiencing (such as the US steel industry being uncompetitive) or being more productive.

Short-Termism

Most political systems allow for governments to be in power for relatively short periods of time. As a result there is always an incentive for politicians to respond more to projects that yield short-term benefits which will maximize their re-election potential, rather than long-term projects that might be economically more efficient but where the benefits might not be realized for some time.

The fallout of the financial crisis of 2007–2009 has highlighted just how many governments across Europe relied on debt financing for government expenditure programmes. The current benefits to countries of these programmes (jobs in public sector activity, spending on the provision of major sporting events such as the Olympics, the World Cup and European Championships, for example) have provided short-term benefits to the population in those countries, but the longer-term necessity to finance this debt by having to increase taxation and impose austerity programmes to help reduce the debt, imposes a significant cost on society as a whole. Once again the value of the benefits is massively outweighed by the costs imposed and represents an economically inefficient allocation.

Public Sector Inefficiency

In the 1980s many governments across the developed world put in place a programme of transferring publicly owned assets to the private sector – so called **privatization**.

privatization the transfer of publicly owned assets to private sector ownership

One of the reasons for this programme was the belief that the public sector cannot run certain types of activity as efficiently as the private sector. This is because the existence of the profit motive in the private sector is a powerful motivator to improve productivity, reduce costs and seek efficiency in production. In the public sector, incentives are different. Managers in public sector operations know that ultimately the taxpayer can bail them out and so the risk inherent in decision making is not the same as that in the private sector where individuals risk their own wealth. In addition, decision makers in the public sector do not gain the same individual benefits and returns that is the case in the private sector if efficiency or productivity is increased and costs cut, and as a result it is more likely that inefficiency will exist.

Cronyism

We have assumed that markets allocate resources on the basis of the interaction between supply and demand and the price mechanism. When government intervenes in the price mechanism through levying taxes and subsidies, regulating business and passing laws which affect the ability of a business to carry out its activities, the market mechanism is distorted and there is the potential for resource allocation to be determined by political rather than economic forces. Where these political forces are influenced by political favours the term **cronyism** is used.

cronyism a situation where the allocation of resources in the market is determined in part by political decision making and favours rather than by economic forces

Cronyism can mean that governments pass laws, institute regulations, levy taxes and impose subsidies which are a response to influence from powerful lobby or special interest groups in return for favours which may increase the welfare of the individual politician or the government as a whole. Such favours might be in the form of a news organization pledging its support for the government, financial donations to the party, personal gifts to politicians or promises of senior posts in management when the individual's political career is over.

Inefficiency in the Tax System

Taxes are a reality in most people's lives. We all have to pay taxes. However, some people or groups are in a position to know enough about the system or have the resources to be able to legitimately avoid paying tax. There are plenty of loopholes in the tax system (which governments create) that individuals can exploit and it can be argued that whilst they are legitimate, it is not in the spirit of the tax system to exploit them. There are also plenty of people who live and work in the so-called 'black' or 'informal economy'. In such cases, these people will earn income, in some cases from criminal activity, which is not declared for tax purposes. This is called tax evasion and is illegal.

The cost to society of the combined activities of tax avoidance and tax evasion is substantial. Because of the very nature of the two, estimates vary on the size of the informal economy but some estimates suggest it could be around 15 per cent of the total tax receipts of some governments. The opportunity cost of such figures is substantial. There are plenty of people who argue that benefit fraud is wrong, but equally there are plenty of people who do not see it as wrong. There are few individuals in society who have not contributed to the informal economy in some way. Most would not see themselves as criminals.

One of the reasons for the existence of tax avoidance and tax evasion might be the design of the tax system. No system is going to be perfect but if people perceive that the system is unfair, there is a greater incentive to find ways around the system or live outside it. In such circumstances, there is a clear case of government failure.

CONCLUSION

The invisible hand is powerful but not omnipotent. A market's equilibrium maximizes the sum of producer and consumer surplus. When the buyers and sellers in the market are the only interested parties, this outcome is efficient from the standpoint of society as a whole. But when there are external effects, such as pollution, evaluating a market outcome requires taking into account the well-being of third parties as well. In this case, the invisible hand of the marketplace may fail to allocate resources efficiently.

In some cases, people can solve the problem of externalities on their own. The Coase theorem suggests that the interested parties can bargain among themselves and agree on an efficient solution. Sometimes, however, an efficient outcome cannot be reached, perhaps because the large number of interested parties makes bargaining difficult.

When people cannot solve the problem of externalities privately, the government often steps in. Yet, even now, society should not abandon market forces entirely. Rather, the government can address the problem by requiring decision makers to bear the full costs of their actions. Pigovian taxes on emissions and pollution permits, for instance, are designed to internalize the externality of pollution. Increasingly, they are being seen as effective policies for those interested in protecting the environment. Market forces, properly redirected, are often the best remedy for market failure.

However, even though government might intervene to correct market failure we also have to account for government failure. Political influence and the incentives of voters, politicians and bureaucrats can often lead to a conflict of interest and as a result the benefits to a small number of people are grossly outweighed by the costs imposed on the majority. In such cases, the attempts to improve market outcomes might lead to further distortions which mean allocations are inefficient.

IN THE NEWS



Gun Crime and Externalities

At the time of writing, the newspapers across the world were reporting the killing of 26 people in Newtown, Connecticut in the United States. The horror of a gunman turning high power weapons on children aged between five and six and their teachers in their school which resulted in 20 children and six adults losing their lives, shocked millions and raised the thorny question of gun control in the United States. It is not only the United States that has witnessed these mass shootings. Has economics got anything to contribute to the debate?

A Market Solution to Gun Control?

The availability of guns in society varies from country to country. In the United States the right to bear arms is considered a fundamental freedom and an issue which millions feel passionate about. Imposing some form of gun control is simply not a topic for negotiation for some. In European countries, different laws

on ownership of guns exist with the requirement of those seeking to own guns to get a licence and to keep their weapons under secure conditions. Regardless of the extent of the law, mass shootings happen. In Europe there have been instances of such events in Switzerland (14 killed in 2001), Germany where 25 people were killed in two separate incidents in 2002 and 2009, both school-based shootings, 12 people killed by

a taxi-driver in Cumbria in the UK in 2011, 69 people at a summer holiday camp in Utoeya, Norway in 2011, six people killed in a shopping mall in the Netherlands in the same year and three people killed in a square in Liege, Belgium, in December 2011.

Looking at these tragic events from the perspective of an economist, the existence of firearms imposes negative externalities on society in the form of the potential for



The costs of gun crime can be substantial but there is a trade-off between the potential for tragedy and the rights of individuals to use weapons for legitimate purposes.

death and injury when they are in the wrong hands. On the other hand, the sport of shooting gives thousands of people considerable pleasure and it is argued in some quarters that there are positive externalities that exist through properly regulated gun ownership in that overall safety is increased if criminals know that they risk being killed because households have firearms to protect themselves. This, it is argued, reduces the incidences of crime which is a benefit to society as a whole. For the purposes of this article, however, we will assume that the negative externalities far outweigh the positive.

Those who produce and sell firearms take into consideration the private costs but they do not take account of the considerable negative externalities that arise from guns falling into the wrong hands. The private market outcome of the quantity and type of guns is significantly higher, therefore, than the socially efficient outcome. Governments could impose taxes on guns to take account of the negative externalities imposed by guns. Should this be a blanket tax on all guns? Some would argue yes, but others would suggest that the tax system could be designed so that those who use guns for legitimate purposes such as

sport and legal hunting would not be affected as much as those who might seek to purchase guns for violent means. For example, high power assault rifles, used by the gunman in Newtown and designed primarily for military purposes, along with ammunition, could be subject to a much higher tax than firearms designed for sports shooting. Along with carefully designed regulation to assess the stability and suitability of those seeking ownership of guns, the potential to arrive at a more socially desirable outcome could be achieved.

Most people would accept that it is impossible to completely eradicate gun crime. Whatever system gun ownership is subjected to there will always be some potential for abuse. Some, however, argue that parts of the gun market are so dangerous that they fall under a category known as noxious markets. A noxious market is one which is considered so morally objectionable that it should not be allowed to exist at all. The idea was put forward by Stanford philosophy professor, Debra Satz. Satz argues that there are moral limits on markets and that some goods should never be for sale. Satz's argument was based around the sale of human organs, specifically kidneys, but can be

extended to some types of firearm. In the case of some types of high powered firearms and ammunition magazines that can enable the user to fire large amounts bullets without having to pause for re-load, the social costs are so high that they come under the classification of a noxious market and, morally, such firearms should simply not be for sale.

Questions

- 1 Can the market for firearms be treated in the same way as the market for pollution (for example)? Why or why not?
- 2 How might the value of the external costs of the firearms market be estimated?
- 3 Use a supply and demand diagram to show the market for firearms. Show the private market outcome and the socially efficient outcome and identify the welfare loss associated with the private market outcome.
- 4 Consider the suggestion to impose taxes on firearms. What sort of design would you recommend and what might be the difficulties associated with such a tax system?
- 5 Should certain types of firearms come under the classification of a noxious market? Why or why not?

SUMMARY

- When a transaction between a buyer and seller directly affects a third party, the effect is called an externality. Negative externalities, such as pollution, cause the socially optimal quantity in a market to be less than the equilibrium quantity. Positive externalities, such as technology spillovers, cause the socially optimal quantity to be greater than the equilibrium quantity.
- Those affected by externalities can sometimes solve the problem privately. For instance, when one business confers an externality on another business, the two

businesses can internalize the externality by merging. Alternatively, the interested parties can solve the problem by negotiating a contract. According to the Coase theorem, if people can bargain without cost, then they can always reach an agreement in which resources are allocated efficiently. In many cases, however, reaching a bargain among the many interested parties is difficult, so the Coase theorem does not apply.

- When private parties cannot adequately deal with external effects, such as pollution, the government often steps in. Sometimes the government prevents socially inefficient activity by regulating behaviour. At other

times it internalizes an externality using Pigovian taxes. Another public policy is to issue permits. For instance, the government could protect the environment by issuing a limited number of pollution permits. The end result of this policy is largely the same as imposing Pigovian taxes on polluters.

- Government intervention to correct market failure might be subject to its own failures. This is because minority groups are able to exercise political power to influence decision making of politicians and bureaucrats to gain benefits which might be outweighed by the costs imposed on the majority.

QUESTIONS FOR REVIEW

- 1 Using examples, explain three sources of market failure.
- 2 Give an example of a negative externality and an example of a positive externality.
- 3 Using an appropriate example, explain the difference between private and social costs and private and social benefits.
- 4 Use a supply and demand diagram to explain the effect of a negative externality in production.
- 5 List some of the ways that the problems caused by externalities can be solved without government intervention.
- 6 Using a demand and supply diagram, represent the market for road use. What is the private market outcome and what is likely to be the socially efficient outcome? Explain and refer to your diagram to help support your answer.
- 7 Imagine that you are a non-smoker sharing a room with a smoker. According to the Coase theorem, what determines whether your room-mate smokes in the room? Is this outcome efficient? How do you and your room-mate reach this solution?
- 8 Use an example to explain the idea of a positional externality.
- 9 How would an extension of property rights help reduce the instances of market failure?
- 10 Explain how government intervention in markets might not improve market outcomes.

PROBLEMS AND APPLICATIONS

- 1 Do you agree with the following statements? Why or why not?
 - a. 'The benefits of Pigovian taxes as a way to reduce pollution have to be weighed against the deadweight losses that these taxes cause.'
 - b. 'When deciding whether to levy a Pigovian tax on consumers or producers, the government should be careful to levy the tax on the side of the market generating the externality.'
- 2 Consider the market for fire extinguishers.
 - a. Why might fire extinguishers exhibit positive externalities?
 - b. Draw a graph of the market for fire extinguishers, labelling the demand curve, the social value curve, the supply curve and the social cost curve.
 - c. Indicate the market equilibrium level of output and the efficient level of output. Give an intuitive explanation for why these quantities differ.
 - d. If the external benefit is €10 per extinguisher, describe a government policy that would result in the efficient outcome.
- 3 In many countries, contributions to charitable organizations are deductible from income tax. In what way does this government policy encourage private solutions to externalities?
- 4 It is rumoured that the Swiss government subsidizes cattle farming, and that the subsidy is larger in areas with more tourist attractions. Can you think of a reason why this policy might be efficient?
- 5 Consider the market for train travel. At certain times of the day, trains are extremely crowded going to major towns and cities but at other times of the day carriages are virtually empty.
 - a. Is this an example of market failure?
 - b. What externalities exist on the train system if the situation described persists?
 - c. Can you think of a way in which the government might step in to make the market outcome more efficient?
- 6 Greater consumption of alcohol leads to more motor vehicle accidents and, thus, imposes costs on people who do not drink and drive.

- a. Illustrate the market for alcohol, labelling the demand curve, the social value curve, the supply curve, the social cost curve, the market equilibrium level of output and the efficient level of output.
 - b. On your graph, shade the area corresponding to the deadweight loss of the market equilibrium. (Hint: the deadweight loss occurs because some units of alcohol are consumed for which the social cost exceeds the social value.) Explain.
- 7** Many observers believe that the levels of pollution in our economy are too high.
- a. If society wishes to reduce overall pollution by a certain amount, why is it efficient to have different amounts of reduction at different firms?
 - b. Command-and-control approaches often rely on uniform reductions among firms. Why are these approaches generally unable to target the firms that should undertake bigger reductions?
 - c. Economists argue that appropriate Pigovian taxes or tradable pollution rights will result in efficient pollution reduction. How do these approaches target the firms that should undertake bigger reductions?
- 8** The Pristine River (or the 'Blue Pristine', as it is affectionately known) has two polluting firms on its banks. European Industrial and Creative Chemicals each dump 100 tonnes of effluent into the river each year. The cost of reducing effluent emissions per tonne equals €10 for European Industrial and €100 for Creative. The government wants to reduce overall pollution from 200 tonnes to 50 tonnes per year.
- a. If the government knew the cost of reduction for each firm, what reductions would it impose to reach its overall goal? What would be the cost to each firm and the total cost to the firms together?
 - b. In a more typical situation, the government would not know the cost of pollution reduction at each firm. If the government decided to reach its overall goal by imposing uniform reductions on the firms, calculate the reduction made by each firm, the cost to each firm and the total cost to the firms together.
- c. Compare the total cost of pollution reduction in parts (a) and (b). If the government does not know the cost of reduction for each firm, is there still some way for it to reduce pollution to 50 tonnes at the total cost you calculated in part (a)? Explain.
 - d. Assume that the chief executive officers of European Industrial and Creative Chemicals have considerable influence with politicians and civil servants from the country's Department of Industry. How might market outcomes be influenced as a consequence?
- 9** Suppose that the government decides to issue tradable permits for a certain form of pollution.
- a. Does it matter for economic efficiency whether the government distributes or auctions the permits? Does it matter in any other ways?
 - b. If the government chooses to distribute the permits, does the allocation of permits among firms matter for efficiency? Does it matter in any other ways?
- 10** Some people argue that the primary cause of global warming is carbon dioxide, which enters the atmosphere in varying amounts from different countries but is distributed equally around the globe within a year. In order to solve this problem, some economists have argued that carbon dioxide emissions should be reduced in countries where the costs are least, with the countries that bear that burden being compensated by the rest of the world.
- a. Why is international cooperation necessary to reach an efficient outcome?
 - b. Is it possible to devise a compensation scheme such that all countries would be better off than under a system of uniform emission reductions? Explain.
 - c. Explain how the concepts of rent seeking and cronyism might be used by critics of climate change theory to argue for different approaches to public policy.

12 INFORMATION AND BEHAVIOURAL ECONOMICS

Economics is a study of the choices that people make and the resulting interactions they have with one another. This study has many facets, as we have seen in the preceding chapters. Yet it would be a mistake to think that all the facets we have seen make up a finished jewel: perfect and unchanging. Like all scientists, economists are always on the lookout for new information which can shed light on existing theories, new areas to study and new phenomena to explain.

Behavioural economics is one area where economists are exploring new boundaries of understanding. This field brings some of the insights from psychology into the study of economic issues. It offers a view of human behaviour that is more subtle and complex than that found in conventional economic theory, but this view may also be more accurate in explaining how humans make choices when compared to the traditional assumptions of rationality and *homoeconomicus*.

Much of the impetus for behavioural views of economics come from other disciplines, notably psychology. Indeed, two of the major thinkers in the field, Daniel Kahneman and Amos Tversky, are psychologists. Kahneman became the first psychologist to win the Nobel Prize for Economics, a prize Kahneman insists would have been shared with his friend and fellow researcher, Amos Tversky, had he not died at the relatively young age of 59 in 1996.

PRINCIPAL AND AGENT

To begin our look at behavioural economics we look at the principal-agent problem. The two parties to an economic decision can be referred to as principal and agent. A **principal** is a person or organization for whom another person called the agent is performing some act. The principal can be seen as the client. An **agent** is a person or organization who is performing an act or acting on behalf of another person or organization (the principal). An agent can be seen as an advisor. The agent invariably has some information which is not known to the principal. In most cases, the principal is using the agent to act for them and to bring about the desired exchange. Because the principal-agent relationship relies on information you will sometimes see this covered under the heading of the *economics of information* or *information economics*.

principal a person for whom another person, called the agent, is performing some act
agent a person who is performing an act for another person, called the principal

The relationship between the principal and agent is a key one and has been of increasing interest in recent years following the financial crisis of 2007–2008. It is often referred to as the principal-agent problem; why is the relationship a problem?

To explain, consider this example. I want to go on holiday and decide to visit a travel agent. I have to rely on the information given to me in making my decision on my holiday. If the travel agent tells me that hotel x is excellent, has superb views, is quiet, peaceful and romantic with excellent food and service, and is valued highly by other customers, how do I know that this is really the case? I might have suspicions that the hotel may be in the middle of a building site, be overrun by noisy children, have less than average

food, poor service and poor quality rooms. I could check the accuracy of the information given to me by the travel agent by going away and looking at a review site on the Internet or checking a satellite image of the area in order to verify the information given to me by the travel agent. But that involves me in additional cost. I might presume that the agent must be giving me accurate information because if they did not they would lose my custom and I could share my experience on other review sites on the Web. The wish to get repeat custom and the profit motive might be a sufficient monitoring device to ensure that the travel agent is acting in my best interest in the advice they are giving me but there may be other motivations that I am unaware of which is driving the advice they are giving. Quite simply, I cannot be sure that the interests of the agent are sufficiently aligned with my interests as the principal. At the heart of the principal-agent problem is asymmetric information.

ASYMMETRIC INFORMATION

Many times in life, one individual, a business or an organization knows more about something than another. In economics, the different access to information of buyers and sellers or any two people is called **asymmetric information**. (Something which is symmetric is identical on both sides – when it is asymmetric, one side is different to another).

asymmetric information where two parties have access to different information

Examples abound. A worker knows more than his employer about how much effort he puts into his job. A seller of a used car knows more than the buyer about the car's condition. The first is an example of a *hidden action*, whereas the second is an example of a *hidden characteristic*. In each case, the party in the dark (the employer, the car buyer) would like to know the relevant information, but the informed party (the worker, the car seller) may have an incentive to conceal it. This is another aspect of the principal-agent problem where invariably the agent has access to information which may not be shared with the agent. It can, however, also be the case that the principal has an incentive to hide information from the agent.

Because asymmetric information is so prevalent, economists have devoted much effort in recent decades to studying its effects. And, indeed, the 2001 Nobel Prize in economics was awarded to three economists (George Akerlof, Michael Spence and Joseph Stiglitz) for their pioneering work on this topic. Let's discuss some of the insights that this research has revealed.

Hidden Actions and Moral Hazard

Moral Hazard **Moral hazard** is a problem that arises when the agent is performing some task on behalf of the principal. In many cases the principal is not able to monitor the behaviour of the agent. This might be because the agent has specific expertise and the principal does not have the knowledge to monitor the agent's behaviour, and even if the principal explained it is not certain the principal could be sure that what they were being told was accurate. In the example of the travel agent above, it is possible to do some research to check the information given by the agent (at additional cost to the principal) but this is not always possible. If the principal cannot perfectly monitor the agent's behaviour, the agent might have an incentive to undertake less effort than the principal considers desirable and as a result the agent may not be fully responsible for the consequences of their actions. The phrase *moral hazard* refers to the risk, or 'hazard', of inappropriate or otherwise 'immoral' behaviour by the agent.

moral hazard the tendency of a person who is imperfectly monitored to engage in dishonest or otherwise undesirable behaviour

The employment relationship is an example where moral hazard occurs. In this situation the employer is the principal and the worker is the agent. The moral-hazard problem is the temptation of imperfectly monitored workers to shirk their responsibilities. Employers can respond to this problem in various ways:

- *Better monitoring.* Parents hiring nannies or au pairs have been known to plant hidden video cameras in their homes to record the individual's behaviour when the parents are away. The aim is to catch irresponsible behaviour.
- *High wages.* Some employers may choose to pay their workers a wage above the level that equilibrates supply and demand in the labour market. A worker who earns an above-equilibrium wage is less likely to shirk, because if he is caught and fired, he might not be able to find another high-paying job. Paying higher wages than equilibrium is called *efficiency wages*.
- *Delayed payment.* Firms can delay part of a worker's compensation, so if the worker is caught shirking and is fired, he suffers a larger penalty. One example of delayed compensation is the year-end bonus. Similarly, a firm may choose to pay its workers more, later in their lives. Thus, the wage rises that workers get as they age may reflect not just the benefits of experience but also a response to moral hazard.

These various mechanisms to reduce the problem of moral hazard need not be used alone. Employers can use a combination of them.

Beyond the workplace there are many other examples of moral hazard. Individuals with insurance cover, be it fire, motor vehicle or medical insurance, may behave differently as a result of having that cover. A motorist, for example, might drive more recklessly in the knowledge that in the event of an accident the cost will be met primarily by the insurance company. Similarly, families choosing to live near a river may benefit from the scenic views but the increased risk of flooding imposes a cost to the insurance company and the government in the event of a serious flood.

Moral hazard has also been much discussed in the wake of the financial crisis: the behaviour of individuals in banks, particularly the investment banking arms of banks. The risks that were being taken by some dealers were considerable. As the problems came to light the potential for the banking system to collapse was considerable and the effect on the wider economy meant that governments and international institutions stepped in to rescue banks. The phrase 'too big to fail' was widely used and it can be argued that many in the banking industry knew this. As a result, dealers and their managers were more willing to take risks that might be considered reckless, even immoral, because they knew that they would not have to face the full consequences of their decision making and actions.

Adverse Selection Moral hazard can lead to **adverse selection**. This means that the market process may end up with 'bad' outcomes because of asymmetric information. Adverse selection occurs when the buyer (principal) knows more about their situation than the seller (agent). The seller knows this and would rather avoid having to do business with these sellers and so might be tempted to charge a higher price as a result. Those buyers who do not fall into the category of being those to be avoided, might be put off buying the good because it is too high a price – the very people who the seller would rather do business with.

adverse selection where a principal knows more about their situation than the agent, leading to the agent preferring not to do business with the principal

Adverse selection is a feature of banking, finance and insurance industries. A bank, for example, may set rules and regulations for its accounts which may lead to some customers, who are not very profitable to the bank, adversely selecting the bank – customers the bank would rather not have. In insurance, the person seeking insurance cover has more information about his or her situation than the insurer. A person who knows they are high risk will look to buy insurance but not necessarily divulge the extent of the risk they pose to the insurance company. The insurance company would prefer not to have to take on these high-risk buyers. It may be difficult for the insurance company to distinguish between its high-risk and low-risk customers; many insurance companies use sophisticated statistical devices to try and do this.

The insurance company would rather take on the low-risk customers than the high-risk ones but high-risk customers adversely select the insurance company. Because of this all seekers of insurance may have to pay higher premiums.

In the wake of the financial crisis, it became clear that some investment banks were putting very risky assets into financial products, and clients buying these products did not know the full extent of the risk they were buying. In this case clients (the principals) were dealing with suppliers (agents) who they would have been better off not dealing with.

Many regulations are aimed at addressing the problem: an insurance company may require homeowners to buy smoke detectors, or pay higher premiums if there is a history of reckless driving (or even refuse to provide insurance cover to the individual), the government may prohibit building homes on land with a high risk of flooding and new regulations may be introduced to curb the behaviour of banks. But the insurance company does not have perfect information about how cautious homeowners are, the government does not have perfect information about the risk that families undertake when choosing where to live and regulators do not know fully the risks that bankers take in investment decisions. As a result, the problem of moral hazard persists.

Hidden Characteristics: Adverse Selection and the Lemons Problem

The Market for Used Cars The classic example of adverse selection is the market for used cars. Sellers of used cars know their vehicles' defects while buyers often do not. Because owners of the worst cars are more likely to sell them than are the owners of the best cars, buyers are apprehensive about getting a poor car. If you are unlucky enough to buy a poor car, then we might say that you have bought a 'lemon'. This was the term used by Nobel Prize-winner George Akerlof in his much cited research article, 'The Market for Lemons' (see Akerlof, G. (1970) 'The Market for Lemons: Quality, Uncertainty and the Market Mechanism'. *Quarterly Journal of Economics*, 84:488–500). His co-prize winners in 2001, Joseph Stiglitz and Michael Spence, also used the term in the context of asymmetric information; it comes from the old-fashioned fruit or gambling machines where three wheels spin and come to rest indicating a picture of a fruit that determines the payout; traditionally, a lemon was bad luck, paying out nothing.

Akerlof used the market for used cars as the basis for his explanation. In this market, the seller has information about the car that the buyer does not have. The seller might know the history of the car and details about how it was driven, whether it has been involved in any accidents and so on. The buyer only has the word of the seller on which to base her decision. Few car buyers have the expertise to be able to conduct a thorough check of a vehicle and be able to verify the seller's claims. This is a clear case of asymmetric information.

Akerlof suggested that there would be two types of cars in the market: good cars, which we will refer to as oranges, and bad cars, which he called 'lemons'. If I go to buy a second-hand car from a dealer, I do not know whether the dealer is selling me an orange or a lemon. I might be willing to pay a reasonable price to buy an orange but I am clearly not willing to pay that same price to buy a lemon. The seller, however, knows whether the car they are trying to sell me is a lemon or an orange. If the seller presents me with a vehicle which he claims is a good quality car, as a buyer, I have to consider whether it is an orange or a lemon. In effect, there is a probability of 0.5 that it will be one or the other.

Given that I have imperfect information, I simply do not know whether the car I am buying is an orange or a lemon. To get an orange, I might have to pay €10,000. The seller on the other hand, who has a lemon for sale, will be prepared to accept almost anything to get rid of it. Let's say that he is willing to accept €4,000 for it. If I offer the seller €4,000 for the car he is trying to sell me and it is a sure-fire orange, he will laugh me out of the showroom. As the buyer, however, I am not sure whether I want to offer €10,000 for the possibility of buying a lemon.

Akerlof also raised the issue about why anyone would want to sell a good car in the first place. If an individual owns a good car, it is highly unlikely that he would get the true value of the car paid to him. Logic would suggest that the only reason he would want to sell his existing car and buy another is because he was looking to replace a sub-standard vehicle with a better quality one. If everyone does this would the market exist at all?

The result is that the market tends to be dominated by lemons. The seller is not willing to sell oranges for less than €10,000 but the buyer is unwilling to pay that much because of the possibility of being sold a lemon.

The conclusion that Akerlof came to was that the market in this type of scenario would only see low quality goods traded.

As a result of this information asymmetry, many people avoid buying vehicles in the used car market. This lemons problem can explain why a used car only a few weeks old sells for thousands of euros less than a new car of the same type. A buyer of the used car might surmise that the seller is getting rid of the car quickly because the seller knows something about it that the buyer does not.

The Labour Market A second example of adverse selection occurs in the labour market. According to efficiency wage theory, workers vary in their abilities, and they may know their own abilities better than do the firms that hire them. When a firm cuts the wage it pays, the more talented workers are more likely to quit, knowing they are better able to find other employment. Conversely, a firm may choose to pay an above-equilibrium wage to attract a better mix of workers.

Or suppose that a firm is not doing so well and needs to cut the wage bill. It can do this either by reducing wages or by keeping wages where they are and laying off workers at random for a few weeks. If it cuts wages, the very best workers will quit, because they know they will be able to find a better job elsewhere. Of course, the better workers who are randomly selected when the firm chooses instead to impose layoffs, may also choose to quit and find a steadier job elsewhere. But in this case only *some* of the best workers quit (since not all of them are laid off because workers were chosen randomly) while if the firm cuts wages *all* of the best workers will quit.

The Insurance Market A third example of adverse selection occurs in markets for insurance. For example, buyers of health insurance know more about their own health problems than do insurance companies. Because people with greater hidden health problems are more likely to buy health insurance than are other people, the price of health insurance reflects the costs of a sicker-than-average person. As a result, people in average health may be discouraged from buying health insurance by the high price.

When markets suffer from adverse selection, the invisible hand does not necessarily work its magic. In the used car market, owners of good cars may choose to keep them rather than sell them at the low price that sceptical buyers are willing to pay. In the labour market, wages may be stuck above the level that balances supply and demand, resulting in unemployment. In insurance markets, buyers with low risk may choose to remain uninsured, because the policies they are offered fail to reflect their true characteristics. Advocates of government-provided health insurance sometimes point to the problem of adverse selection as one reason not to trust the private market to provide the right amount of health insurance on its own.

Signalling to Convey Private Information

Although asymmetric information is sometimes a motivation for public policy, it also motivates some individual behaviour that otherwise might be hard to explain. Markets respond to problems of asymmetric information in many ways. One of them is **signalling**, which refers to actions taken by an informed party for the sole purpose of credibly revealing his private information.

signalling an action taken by an informed party to reveal private information to an uninformed party

Firms may spend money on advertising to signal to potential customers that they have high-quality products. Students may earn university degrees in order to signal to potential employers that they are high-ability individuals. The signalling theory of education asserts that education increases a person's productivity, rather than merely conveying information about innate talent. These two examples of signalling (advertising, education) may seem very different, but below the surface they are much the same: in both cases, the informed party (the firm, the student) is using the signal to convince the uninformed party (the customer, the employer) that the informed party is offering something of high quality.

What does it take for an action to be an effective signal? Obviously, it must be costly. If a signal were free, everyone would use it, and it would convey no information. For the same reason, there is another

requirement: the signal must be less costly, or more beneficial, to the person with the higher-quality product. Otherwise, everyone would have the same incentive to use the signal, and the signal would reveal nothing.

Consider again our two examples. In the advertising case, a firm with a good product reaps a larger benefit from advertising because customers who try the product once are more likely to become repeat customers. Thus, it is rational for the firm with the good product to pay for the cost of the signal (advertising), and it is rational for the customer to use the signal as a piece of information about the product's quality. In the education case, a talented person can get through school more easily than a less talented one. Thus, it is rational for the talented person to pay for the cost of the signal (education), and it is rational for the employer to use the signal as a piece of information about the person's talent.

CASE STUDY The Deadweight Loss of Christmas

For millions of people around the world, 25 December is a day for exchanging gifts. Using the standard economic model to explain gift-giving presents problems because the assumption of the self-interest of the purchaser is distorted by the fact that the purchaser is buying something not to satisfy their own preferences but anticipating the preferences of the recipient. Joel Waldfogel reviewed this issue in a 1993 paper (Waldfogel, J. (1993) 'The Deadweight Loss of Christmas'. *The American Economic Review*, December, 83:5:1328–1336). Waldfogel began with the premise that the optimal outcome for a gift-giver would be to replicate the choices the recipient would have made if they had spent the same amount of money the gift-giver spent. The difference between the two would be the deadweight loss of gift-giving.

Waldfogel's research suggested that the size of the deadweight loss could be between a tenth and a third of the value of the gift. Gift-giving can be inefficient when the utility of the recipient could have been higher if she had spent the same amount of money that the gift-giver spent himself. The size of the deadweight loss is dependent on the extent to which the gift-giver can replicate the preferences of the recipient – and this obviously depends on the extent of the information asymmetry between them. If the gift-giver knows the recipient extremely well (i.e. has perfect information about the recipient's preferences) then he can replicate that person's preferences and the gift will yield value above its cost. However, the existence of information asymmetry is likely to mean that there will be some deadweight loss associated with gift-giving.

What is the best sort of gift to give? The gift-giver has to estimate the utility the recipient will receive from the gift. If he gives cash the recipient can use that cash to maximize her preferences, but any other gift may not match the utility the recipient could have achieved if she had spent the money herself. The more ignorant the gift-giver is of the preferences of the recipient the greater the deadweight loss. It can be argued, therefore, that giving cash is always the best gift to give unless the giver knows the preferences of the recipient extremely well.

This brings us to the type of giver; is one type of giver likely to have more knowledge about the preferences of the recipient than another? Waldfogel's research suggested that aunts/uncles and grandparents' gifts are most likely to be associated with higher deadweight losses, whereas those from friends and significant others have the lowest.

Giving gifts is fraught with difficulty caused by asymmetric information. The gift giver in this instance looks like they did not fully understand the preferences of the recipient. There is deadweight loss!



The world is replete with instances of signalling. Magazine advertisements sometimes include the phrase ‘as seen on TV’. Why does a firm selling a product in a magazine choose to stress this fact? One possibility is that the firm is trying to convey its willingness to pay for an expensive signal (a spot on television) in the hope that you will infer that its product is of high quality. For the same reason, graduates of elite universities are always sure to put that fact on their *curriculum vitae*.

Screening to Induce Information Revelation

When an uninformed party takes actions to induce the informed party to reveal private information, the phenomenon is called **screening**.

screening an action taken by an uninformed party to induce an informed party to reveal information

Some screening is common sense. A person buying a used car may ask that it be checked by a car mechanic before the sale. A seller who refuses this request reveals his private information that the car is a lemon. The buyer may decide to offer a lower price or to look for another car.

Other examples of screening are more subtle. For example, consider a firm that sells car insurance. The firm would like to charge a low premium to safe drivers and a high premium to risky drivers. But how can it tell them apart? Drivers know whether they are safe or risky, but the risky ones won’t admit to it. A driver’s history is one piece of information (which insurance companies in fact use), but, because of the intrinsic randomness of car accidents, history is an imperfect indicator of future risks.

The insurance company might be able to sort out the two kinds of drivers by offering different insurance policies that would induce them to separate themselves. One policy would have a high premium and cover the full cost of any accidents that occur. Another policy would have low premiums but would have, say, a €1,000 excess. (That is, the driver would be responsible for the first €1,000 of damage, and the insurance company would cover the remaining risk.) Notice that the excess is more of a burden for risky drivers because they are more likely to have an accident. Thus, with a large enough excess, the low-premium policy with an excess would attract the safe drivers, while the high-premium policy without an excess would attract the risky drivers. Faced with these two policies, the two kinds of drivers would reveal their private information by choosing different insurance policies.

Asymmetric Information and Public Policy

We have examined two kinds of asymmetric information – moral hazard and adverse selection – and we have seen how individuals may respond to the problem with signalling or screening. Now let’s consider what the study of asymmetric information suggests about the proper scope of public policy.

The tension between market success and market failure is central in microeconomics. We have learned that the equilibrium of supply and demand is efficient in the sense that it maximizes the total surplus that society can obtain in a market. Adam Smith’s invisible hand seemed to reign supreme. This conclusion is tempered by the problem of public goods and externalities. We will see more examples of market failure when we look in more detail at firms’ behaviour when the assumption of perfect competition is dropped. These examples of market failure showed that government can sometimes improve market outcomes, notwithstanding the potential for government failure.

The study of asymmetric information gives us new reason to be wary of markets. When some people know more than others, the market may fail to put resources to their best use. People with high-quality used cars may have trouble selling them because buyers will be afraid of getting a lemon. People with few health problems may have trouble getting low-cost health insurance because insurance companies lump them together with those who have significant (but hidden) health problems.

Although asymmetric information may call for government action in some cases, three facts complicate the issue. First, as we have seen, the private market can sometimes deal with information asymmetries on its own using a combination of signalling and screening. Second, the government rarely has more

information than the private parties. Even if the market's allocation of resources is not first-best, it may be second-best. That is, when there are information asymmetries, policymakers may find it hard to improve upon the market's admittedly imperfect outcome. Third, the government is itself an imperfect institution, as we have seen in our analysis of government failure.

SELF TEST A person who buys a life insurance policy pays a certain amount per year and receives for his family a much larger payment in the event of his death. Would you expect buyers of life insurance to have higher or lower death rates than the average person? How might this be an example of moral hazard? Of adverse selection? How might a life insurance company deal with these problems?

DEVIATIONS FROM THE STANDARD ECONOMIC MODEL

Economics is a study of human behaviour, but it is not the only field that can make that claim. The social science of psychology also sheds light on the choices that people make in their lives. The fields of economics and psychology usually proceed independently, in part because they address a different range of questions. Behavioural economics makes use of basic psychological insights to explain human behaviour especially when faced with making decisions and when faced with choices.

In Chapter 5 we introduced the standard economic model (SEM) and looked at some behavioural approaches to explaining consumer behaviour, which cast doubt on the assumption that people always behave rationally.

People Aren't Always Rational

Economic theory is populated by a particular species of organism, sometimes called *Homoeconomicus*. Two prominent behavioural economists from the University of Chicago, Richard Thaler and Cass Sunstein, refer to them as '*econs*'. Members of this species are always rational. As managers of firms, they maximize profits. As consumers, they maximize utility (or, equivalently, pick the point on the highest indifference curve). Given the constraints they face, they rationally weigh all the costs and benefits and always choose the best possible course of action.

Real people, however, are *Homo sapiens*. Although in many ways they resemble the rational, calculating people assumed in economic theory, they are far more complex. They can be forgetful, impulsive, confused, emotional and short-sighted. These imperfections of human reasoning are the bread-and-butter of psychologists.

Herbert Simon, one of the first social scientists to work at the boundary of economics and psychology, suggested that humans should be viewed not as rational maximizers but as **satisficers**. Rather than always choosing the best course of action, they make decisions that are merely good enough.

satisficers those who make decisions based on securing a satisfactory rather than optimal outcome

We saw in Chapter 5 how humans might be considered as only 'near rational' or that they exhibit *bounded rationality*. We also saw how humans make systematic mistakes or errors in decision making under different conditions. What follows are some further findings which have an effect on consumer decision making.

Mental Accounting You may know people who have tins which they use to collect money which will be used for a particular purpose – electricity, gas, holidays and so on. This is an example of how people have a tendency to separate money into different accounts based on different criteria. Having these accounts

might provide the individual with some comfort but they might also be completely irrational. For example, if an individual diverts some portion of her income into a savings account each month to save for a holiday whilst at the same time only paying off the minimum amount on her credit card, this would be deemed highly irrational. The diverted money would be better spent on reducing the debt, which is also accruing interest over time – the longer the debt is deferred the larger the debt will be. The fact that people do not do this suggests that they attach a subjective value to each ‘account’ which is not a logical and rational allocation of their funds.

Similarly, when people receive funds from unexpected sources, such as a tax rebate or a bonus payment from work, they tend to view it and spend it differently to their ‘normal income’. This is also an example of irrational behaviour; regardless of the origin of the money, it is still money and can be allocated in exactly the same way to satisfy preferences.

Herd Mentality There are occasions when people make decisions which follow those of a much larger group – sometimes this group may not be tangible but for some reason individuals are persuaded by the apparent power of the group. For example, in periods where house or stock prices are rising there may be a tendency for individuals to make decisions on the purchase of these assets which are at odds with their ‘true value’. Indeed, this is how asset price bubbles arise. Individuals may make decisions based on herd behaviour because of reasons of social conformity (humans have a tendency to want to ‘belong’ to a group and to reflect the behaviour of the group) and because there may be a sense of ‘if so many people are making these decisions they can’t all be wrong’, and that other people may be more informed.

Prospect Theory Imagine that someone offered you the prospect of winning €200 but then a day later losing €100, or of winning €100. Which would you choose? Research suggests that more people would choose the second option. Closer inspection of the choices offered reveal that they are both the same in terms of the net gain to individuals – the net gain in each case is €100. So why do more people choose the second option if the net gain is the same?

Kahneman and Tversky suggested the reason was because people do not value gains and losses in the same way. Losses appear to be attached with more emotion than do gains. **Prospect theory** suggests that when presented with different prospects (outcomes) from a transaction or an exchange, people will value the losses and gains differently even if the value of each is the same. This helps explain, for example, why some people are willing to drive 25km to save €10 on a €50 item but would not be willing to drive the same distance to save the same amount of money (€10) on a €500 item even though the saving is the same.

prospect theory a theory that suggests people attach different values to gains and losses and do so in relation to some reference point

This insight has important implications for financial decisions where the risk of making losses in relation to the gains that could be made are important. In such situations, decision makers might exhibit risk aversion to the potential for making losses but undervalue the potential gains that could be made. This is partly because humans attach reference points to decision making which mean we place values on changes rather than on absolute magnitudes. For example, if we are in a room which is heated to 35 degrees Celsius and then walk into a room which is heated to 25 degrees Celsius we would suggest that the room is ‘cold’ when at other times 25 degrees Celsius is considered by most people to be ‘warm’.

When evaluating gains and losses from decision making, therefore, we react to some reference point. This is possibly why humans place such a value on a loss rather than a gain as the gain does not provide the same reference point as to a loss. This also helps explain why the owner of something tends to place a higher value on it than anyone else – the **endowment effect**.

endowment effect where the value placed on something owned is greater than on an identical item not owned

If we are putting up a house for sale, for example, not only are we pricing the asset value of the house in terms of the bricks and mortar components but there will also be a considerable emotional investment in a property we own and when we come to sell it the value of that emotional investment will be priced into the sale – it is the value we place on losing all those emotional experiences. For the buyer who has not got these emotional ties to the property, the value will be far less. Similar experiences exist when people access tickets for key sporting or musical events. Once the ticket is ‘owned’ it tends to be valued far more highly because the owner can now begin to articulate in their mind the value of the loss they will experience if they give up the ticket.

The way losses and gains are presented, therefore, can have an effect on the way in which we make choices and decisions and the way in which businesses can exploit human psychology. This is linked to the section on framing outlined in Chapter 5.

People Care about Fairness

Another insight about human behaviour is best illustrated with an experiment called the *ultimatum game*. The game works like this: two volunteers (who are otherwise strangers to each other) are told that they are going to play a game and could win a total of €100. Before they play, they learn the rules. The game begins with a flip of a coin, which is used to assign the volunteers to the roles of player A and player B. Player A's job is to propose a division of the €100 prize (in whole euros) between himself and the other player. After player A makes his proposal, player B decides whether to accept or reject it. If he accepts it, both players are paid according to the proposal. If player B rejects the proposal, both players walk away with nothing. In either case, the game then ends.

Before proceeding, stop and think about what you would do in this situation. If you were player A, what division of the €100 would you propose? If you were player B, what proposals would you accept?

Conventional economic theory assumes in this situation that people are rational wealth-maximizers. This assumption leads to a simple prediction: player A should propose that he gets €99 and player B gets €1, and player B should accept the proposal. After all, once the proposal is made, player B is better off accepting it as long as he gets something out of it (remember, people are assumed to act at the margin – a €1 euro gain is better than nothing). Moreover, because player A knows that accepting the proposal is in player B's interest, player A has no reason to offer him more than €1.

Yet when experimental economists ask real people to play the ultimatum game, the results are very different from this prediction. People in the role of player B usually reject proposals that give them only €1 or a similarly small amount. Knowing this, people in the role of player A usually propose giving player B much more than €1. Some people will offer a 50–50 split, but it is more common for player A to propose giving player B an amount such as €30 or €40, keeping the larger share for himself. In this case, player B usually accepts the proposal.

What's going on here? The natural interpretation is that people are driven in part by some innate sense of fairness. A 99–1 split seems so wildly unfair to many people that they reject it, even to their own detriment. By contrast, a 70–30 split is still unfair, but it is not so unfair that it induces people to abandon their normal self-interest.

Throughout our study of household and firm behaviour, the innate sense of fairness has not played any role. But the results of the ultimatum game suggest that perhaps it should. We will see in later chapters that wages are determined by labour supply and labour demand. Some economists have suggested that the perceived fairness of what a firm pays its workers should also enter the picture. Thus, when a firm has an especially profitable year, workers (like player B) may expect to be paid a fair share of the prize, even if the standard equilibrium does not dictate it. The firm (like player A) might well decide to give workers more than the equilibrium wage for fear that the workers might otherwise try to punish the firm with reduced effort, strikes or even vandalism.

To return to the ultimatum game, do you think that a sense of fairness may have its price? If the players were given, say, €1,000 to divide to the nearest hundred, and player A proposed a split of €900 to him and €100 to player B, do you think that player B would be just as likely to reject the proposal as before? What if the prize money were raised to €1 million, to be divided to the nearest €100,000? The answer may depend on framing – behaviour and decision making will be dependent on the way decision problems or choices are framed.

People Are Inconsistent over Time

Imagine some dreary task, such as doing your laundry or tidying your room. Now consider the following questions:

1. Would you prefer (A) to spend 50 minutes doing the task immediately or (B) to spend 60 minutes doing the task tomorrow?
2. Would you prefer (A) to spend 50 minutes doing the task in 90 days or (B) to spend 60 minutes doing the task in 91 days?

When asked questions like these, many people choose B to question 1 and A to question 2. When looking ahead to the future (as in question 2), they minimize the amount of time spent on the dreary task. But faced with the prospect of doing the task immediately (as in question 1), they choose to put it off.

In some ways, this behaviour is not surprising: everyone procrastinates from time to time. But from the standpoint of the theory of rational humans, it is puzzling. Suppose that, in response to question 2, a person chooses to spend 50 minutes in 90 days. Then, when the 90th day arrives, we allow him to change his mind. In effect, he then faces question 1, so he opts for doing the task the next day. But why should the mere passage of time affect the choices he makes?

Many times in life, people make plans for themselves, but then they fail to follow them through. A smoker promises herself that she will quit, but within a few hours of smoking her last cigarette, she craves another and breaks her promise. A person trying to lose weight promises that he will stop eating chocolate bars, but when he gets to the checkout at the supermarket and sees the tempting array of confectionery next to the cash register, the promise is forgotten. In both cases, the desire for instant gratification induces the decision maker to abandon their own past plans.

Some economists believe that the consumption-saving decision is an important instance where people exhibit this inconsistency over time. For many people, spending provides a type of instant gratification. Saving, like passing up the cigarette or the dessert, requires a sacrifice in the present for a reward in the distant future. And just as many smokers wish they could quit and many overweight individuals wish they ate less, many consumers wish they saved more.

An implication of this inconsistency over time is that people should try to find ways to commit their future selves to following through on their plans. A smoker trying to quit may throw away her cigarettes, and a person on a diet may put a lock on the refrigerator and ask someone else to do the shopping. What can a person who saves too little do? He should find some way to lock up his money before he spends it. Some personal pension plans do exactly that. A worker can agree to have some money taken out of his salary payment before he ever sees it. The money is placed in an account and invested on his behalf by the pension company. When he retires, he can use the money to fund a pension, but the money can only be used before retirement with a penalty. This is one reason why people take out pension plans: they protect people from their own desires for instant gratification.

SELF TEST Describe at least three ways in which human decision making differs from that of the rational individual of conventional economic theory.

CONCLUSION

This chapter has examined some of the issues in behavioural explanations of human decision making – behavioural economics. You may have noticed that we have sketched out ideas rather than fully developing them. This is no accident. One reason is that you might study these topics in more detail in advanced courses. Another reason is that these topics remain active areas of research and, therefore, are still being fleshed out.

To see how these topics fit into the broader picture, recall the *Ten Principles of Economics*. One principle states that markets are usually a good way to organize economic activity. Another principle states that governments can sometimes improve market outcomes. As you study economics, you can more fully appreciate the truth of these principles as well as the caveats that go with them. The study of asymmetric

IN THE NEWS



Moral Hazard

In 2012 news emerged that some banks had conspired to influence a key interest rate setting process. Some individuals had to answer to parliamentary committees and answer some uncomfortable questions. Tens of thousands of people around Europe and beyond might have been affected by the rate fixing but progress on identifying the culprits has been slow.

'None of them has gone to jail': man who ran the World Bank calls for bankers to face the music

The Barclays Libor scandal may have shocked the British public, but Joseph Stiglitz saw it coming decades ago. And he's convinced that jailing bankers is the best way to curb market abuses. A towering genius of economics, Stiglitz wrote a series of papers in the 1970s and 1980s explaining how, when some individuals have access to privileged knowledge that others don't, free markets yield bad outcomes for wider society. That insight (known as the theory of 'asymmetric information') won Stiglitz the Nobel Prize for economics in 2001.

Joseph Stiglitz won the Nobel Peace Prize for his work on asymmetric information.



And he has leveraged those credentials relentlessly ever since to batter at the walls of 'free market fundamentalism'. It is a crusade that has taken Stiglitz from the Massachusetts Institute of Technology, to the Clinton White House, to the World Bank, to the Occupy Wall Street camp and, now, to London, to promote his new book *The Price of Inequality*. And kind fortune has engineered it so that Stiglitz's UK trip has coincided with a perfect example of the repellent consequences of asymmetric information.

When traders working for Barclays rigged the Libor interest rate and flogged toxic financial derivatives – using their privileged position in the financial system to make profits at the expense of their customers – they were unwittingly proving Stiglitz right.

'It's a textbook illustration,' Stiglitz said. 'Where there are these asymmetries a lot of these activities are directed at rent seeking [appropriating resources from someone else rather than creating new wealth]. That was one of my original points. It wasn't about productivity, it was taking advantage.'

Yet Stiglitz's interest in the abuses of banks extends beyond the academic. He argues that breaking the economic and political power that has been amassed by the financial sector in recent decades, especially

in the US and UK, is essential if we are to build a more just and prosperous society. The first step, he says, is sending some bankers to jail. 'That ought to change. That means legislation. Banks and others have engaged in rent seeking, creating inequality, ripping off other people, and none of them has gone to jail.'

Next, politicians need to stop spending so much time listening to the financial lobby, which, according to Stiglitz, demonstrates its spectacular economic ignorance whenever it claims that curbs on banks' activities will damage the broader economy.

Questions

- 1 How does the fixing of interest rates by bankers illustrate the principle of asymmetric information?
- 2 What do you think is meant by the phrase 'free market fundamentalism'?
- 3 How might the existence of asymmetries in information lead to rent seeking?
- 4 How would jailing bankers (assuming it could be proved they were guilty) help reduce moral hazard in banking?
- 5 Banks are an integral part of the financial system, one which is a key part of the market mechanism. If free markets allocate resources efficiently, why do you think Stiglitz believes the power of banks should be broken if a more 'just and prosperous society' is to be built?

information should make you more wary of market outcomes and the study of behavioural economics should make you wary of any institution that relies on human decision making – including both the market and the government.

If there is a unifying theme to these topics, it is that life is messy. Information is imperfect, government is imperfect and people are imperfect. Of course, you knew this long before you started studying economics, but economists need to understand these imperfections as precisely as they can if they are to explain, and perhaps even improve, the world around them.

SUMMARY

- In many economic transactions, information is asymmetric. When there are hidden actions, principals may be concerned that agents suffer from the problem of moral hazard. When there are hidden characteristics, buyers may be concerned about the problem of adverse selection among the sellers. Private markets sometimes deal with asymmetric information with signalling and screening.
- The study of psychology and economics reveals that human decision making is more complex than is assumed in conventional economic theory. Human beings make errors in decision making that reflect biases and can also be influenced by the way information is framed and how they value outcomes. People are not always rational; they care about the fairness of economic outcomes (even to their own detriment), and they can be inconsistent over time.

QUESTIONS FOR REVIEW

- 1 Explain the difference between a principal and an agent.
- 2 Why does the existence of asymmetric information mean that economic decisions do not conform to the assumptions of the standard economic model?
- 3 What is the difference between a hidden action and a hidden characteristic?
- 4 What is moral hazard? List three things an employer might do to reduce the severity of this problem.
- 5 What is adverse selection? Give an example of a market in which adverse selection might be a problem.
- 6 Define *signalling* and *screening*, and give an example of each.
- 7 Why can mental accounting lead to irrational decision making?
- 8 Why might herd mentality lead to asset prices rising faster than the fundamental value of the assets?
- 9 Use an example to explain prospect theory.
- 10 Describe the ultimatum game. What outcome from this game would conventional economic theory predict? Do experiments confirm this prediction? Explain.

PROBLEMS AND APPLICATIONS

- 1 In each of the following cases, who is the principal and who is the agent and explain why there might be asymmetry of information.
 - a. A suspicious wife hires a private detective to report on the movements of her husband.
 - b. A car leasing company providing a holiday maker with a rental vehicle.
 - c. A homeowner seeking insurance for her home against flooding.
 - d. An individual making a routine visit to the dentist for a check up.
- 2 In the situations in Question 1 above, identify the hidden actions and the hidden characteristic/s.
- 3 Each of the following situations involves moral hazard. In each case, identify the principal and the agent, and explain why there is asymmetric information. How does the action described reduce the problem of moral hazard?
 - a. Landlords require tenants to pay security deposits.
 - b. Firms compensate top executives with options to buy company shares at a given price in the future.
 - c. Car insurance companies offer discounts to customers who install anti-theft devices in their cars.
- 4 Suppose that the Live-Long-and-Prosper Health Insurance Company charges €5,000 annually for a family insurance policy. The company's president

suggests that the company raise the annual price to €6,000 in order to increase its profits. If the firm followed this suggestion, what economic problem might arise? Would the firm's pool of customers tend to become more or less healthy on average? Would the company's profits necessarily increase?

- 5 A boyfriend can signal to a girlfriend that he loves her by giving an appropriate gift. Do you think saying 'I love you' can also serve as a signal? Why or why not?
- 6 Some AIDS activists believe that health insurance companies should not be allowed to ask applicants if they are infected with the HIV virus that causes AIDS. Would this rule help or hurt those who are HIV-positive? Would it help or hurt those who are not HIV-positive? Would it exacerbate or mitigate the problem of adverse selection in the market for health insurance? Do you think it would increase or decrease the number of people without health insurance? In your opinion, would this be a good policy?
- 7 The government is considering two ways to help the needy: giving them cash, or giving them free meals at soup kitchens. Provide an argument for giving cash. Present an argument, based on asymmetric information, for why the soup kitchen may be better than the cash handout.
- 8 Michael turns up to an interview in a brand new designer suit. What do you think Michael is trying to do? How might the employer find out if Michael represents a good investment as a prospective employee?
- 9 Imagine that you are in charge of health policy in your country. A mutation of the swine flu virus leads to a rapid

increase in the number of people contracting the virus with the expectation that 1,200 people will die as a result. There are medications and inoculation programmes that will help save some lives but it is inevitable that some people will die. Your problem is presenting the news to the population. Your officials present you with the following wordings as part of your speech on the policy:

- a. Our inoculation programme will save the lives of 400 people
- b. We are going to offer a programme to all the population that will have a one-third chance that 1,200 people will live
- c. Unfortunately, despite the best efforts of this government, our inoculation programme will not be enough to save everyone and so I have to tell you that even with the programme 400 people will die
- d. Our inoculation programme will help but I have to be honest with you and tell you that the programme has a one-third chance that no-one will die but a two-thirds chance that 1,200 people will die.

Which option would you choose? Does it matter which one you chose? Why? Why not? How does this highlight prospect theory and the idea of framing?

- 10 Two ice-cream stands are deciding where to locate along a one-mile beach. Each person sitting on the beach buys exactly one ice-cream cone per day from the stand nearest to her. Each ice cream seller wants the maximum number of customers. Where along the beach will the two stands locate?

PART 6

FIRM BEHAVIOUR AND MARKET STRUCTURES

13 FIRMS' PRODUCTION DECISIONS

In Chapter 6 we looked at firms' costs and revenues and the profit maximizing output. We also looked at the derivation of the supply curve. Our analysis was based on the assumption of firms operating in perfectly competitive conditions. In this chapter we are going to look in more detail at a firm's production decisions and then move on in subsequent chapters to see how firms' behaviour changes when the assumptions of perfect competition are dropped.

ISOQUANTS AND ISOCOSTS

One of the issues facing businesses in considering production decisions is to attempt to maximize output or minimize costs but with a constraint of limited factor inputs. This is a further example of a constrained optimization problem. Different firms have different ratios of factor inputs – land, labour and capital – in the production process. This can vary not only between industries but also within industries. For example, some farms are highly land intensive whereas others may be far more capital or labour intensive. Arable farms tend to have a very high proportion of land in comparison to a pig farm and be classed as land intensive. Poultry farms may employ a much larger number of workers than an arable farm and be classed as labour intensive by comparison.

Businesses can utilize their factors of production in different ways to produce any given output so an important question that firms need to address is how to organize its factor inputs to maximize output at minimum cost. The use of isocost and isoquant lines provides a model to help conceptualize the process. The principles are very similar to the model of consumer behaviour we introduced in Chapter 5, which made use of indifference curves and budget constraints. In this model the firm is faced with different combinations of factors which yield the same amount of output (isoquants) and the budget available to pay for those factors of production (isocosts).

Production Isoquants

A **production isoquant** is a function which represents all the possible combinations of factor inputs that can be used to produce a given level of output. For simplicity we are going to assume just two factor inputs: labour and capital. To further focus our thinking, let us assume that the capital in question is a

machine that coats pizzas with a tomato base, then adds the filling and then bakes the pizzas. The number of hours the machine has to operate varies with the amount of pizzas produced. The labour will be the number of person hours needed to mix and produce the dough for the pizza base, feed the machine and then package the finished pizzas.

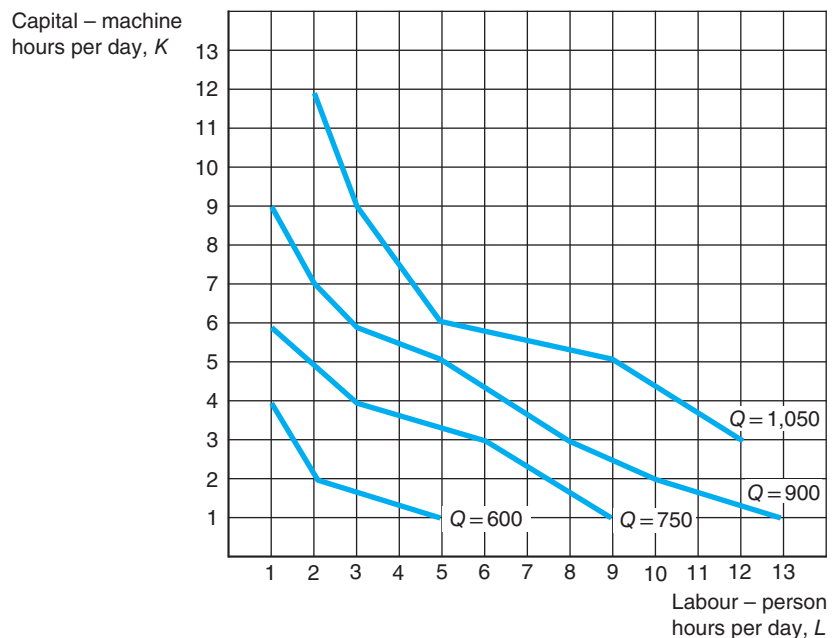
production isoquant a function representing all possible combinations of factor inputs that can be used to produce a given level of output

Figure 13.1 shows a graphical representation of the production isoquants that relate to the combinations of labour and capital that can be used to produce pizzas. An output level of $Q = 600$ could be produced using 5 hours of labour and 1 hour of the machine or 2 hours of labour and 2 hours of the machine. The isoquant line $Q = 600$ connects all the possible combinations of capital and labour which could produce an output of 600 pizzas. Given the level of capital and labour inputs for the pizza factory, a series of isoquants can be drawn for different levels of output. Figure 13.1 shows the isoquants for output levels of $Q = 600$, $Q = 750$, $Q = 900$ and $Q = 1,050$. In theory, the whole of the graphical space could be covered with isoquants all relating to the different levels of possible output.

FIGURE 13.1

Production Isoquants for a Pizza Factory

Given the possibility of employing different amounts of capital and labour, the isoquant map connects together combinations of capital and labour which could be employed to produce different levels of output of pizzas. For an output level $Q = 600$, the machine operating for 4 hours and 1 hour of labour time could produce 600 pizzas per day but so could the combination 2 hours for the machine and 2 hours of labour time. Five hours of the machine and 5 hours of labour could produce an output level $Q = 900$; a combination of 2 hours of the machine and 10 hours of labour could also produce 900 pizzas.



Few businesses will sit down and draw out isoquants in the way we have done here, but the reality is that firms do regularly make decisions about factor combinations in deciding output. Firms will often look at the option of substituting capital for labour by making staff redundant and investing instead in new equipment. Firms may also look at replacing existing machinery for new ones or look for outsourcing opportunities both of which would have an effect on the shape and position of the isoquants.

Substituting one factor for another will incur costs. It may not be easy to be able to substitute one factor for another; machinery may be highly specialized and workers may have skills that machines simply cannot replicate (the ability to make clients feel confident and at ease, for example). The slope of the isoquant

represents the **marginal rate of technical substitution** (MRTS). This is the rate at which one factor input can be substituted for another at a given level of output. Referring to Figure 13.1, take the output level $Q = 1,050$ and a combination of labour and capital at 5 and 6 hours respectively.

marginal rate of technical substitution the rate at which one factor input can be substituted for another at a given level of output

If the owner of the pizza factory considered reducing labour hours by 2 she would have to increase the amount of hours the machine was used by 3 to 9 hours in order to maintain output at 1,050. The MRTS would be given by the ratio of the change in capital to the change in labour, $\Delta K/\Delta L$. The change in capital is from 6 to 9 units and the change in labour is from 5 to 3. The $MRTS = 3/2$ or 1.5. This tells us that the owner has to increase the amount of hours of capital by 1.5 for every 1 hour of labour reduced to maintain production at $Q = 1,050$.

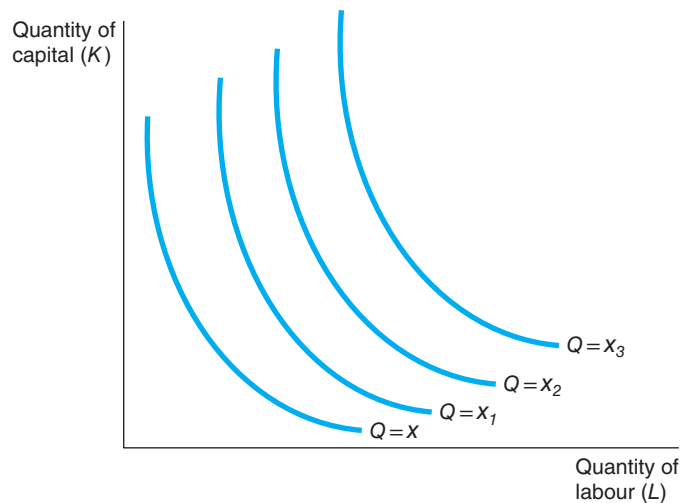
The way we have drawn the isoquants in Figure 13.1 would suggest different MRTS at different points because the slope of each isoquant is different.

It is common to see isoquants drawn as smooth curves, as shown in Figure 13.2. Clearly, in such a situation the use of calculus to calculate the MRTS would be advantageous because it would provide a very much more accurate value at every point on each curve. Using calculus the derivative of capital with respect to labour $\left(\frac{dK}{dL}\right)$ would give us this precise value.

FIGURE 13.2

Production Isoquants

It is common to represent production isoquants as a series of smooth curves representing the different combinations of capital and labour which would be used to produce different levels of output represented in this figure as $Q = x$, $Q = x_1$, $Q = x_2$, etc.



When a firm reduces one unit of a factor and substitutes it for another, unless the factors are perfect substitutes, it is likely that the addition to total output of each successive unit of the factor employed will diminish according to the law of diminishing marginal productivity. At the same time, as less of the other factor is used its marginal product will be higher. This helps explain why isoquants are convex to the origin. If a firm employs a large amount of capital and not much labour, the substitution of one unit of capital for a unit of labour, the marginal product of that extra unit of labour is likely to be high (and the slope of the isoquant relatively steep) but as additional units of labour are substituted the marginal product diminishes and so the slope of the isoquant gradually gets less steep. The MRTS is the ratio of the marginal products of labour and capital:

$$MRTS = \frac{MP_L}{MP_K}$$

Where the

$$MP_L = \frac{\text{change in the quantity of output } (\Delta Q)}{\text{change in the quantity of labour } (\Delta L)} \text{ and}$$

$$MP_K = \frac{\text{change in the quantity of output } (\Delta Q)}{\text{change in the quantity of capital } (\Delta K)}$$

Isocost Lines

Our analysis so far has looked at different combinations of factor inputs to produce given outputs. A business has to take into consideration the obvious fact that factor inputs cost money. Labour has to be paid wages and salaries, and energy to power the machines has to be purchased in addition to the cost of the machines themselves. Firms do not have unlimited funds to purchase factor inputs and so have purchase constraints. Many firms will set budgets for purchasing factor inputs which have to be adhered to.

Isocost lines take the cost of factor inputs into consideration. An **isocost line** shows the different combination of factor inputs which can be purchased with a given budget.

isocost line a line showing the different combination of factor inputs which can be purchased with a given budget

Assume that the price of an hour's operation of the pizza machine is P_k then the cost of capital would be $P_k K$ (the price of capital multiplied by the amount of capital hours used), and the cost of labour is given by $P_L L$ (the price of labour multiplied by the amount of labour hours). Given a budget constraint represented by TC_{KL} we can express the relationship as:

$$P_k K + P_L L = TC_{KL}$$

Now assume that the price of capital to make pizzas is €10 per hour and the price of labour, €6 per hour. Our formula would look like this:

$$10K + 6L = TC_{KL}$$

Using 3 capital hours and 9 hours of labour would cost $10(3) + 6(9) = €84$. Are there other combinations of capital and labour that would produce pizzas at a cost of €84? We can find this out by rearranging the equation to give:

$$€84 = 10K + 6L$$

We can now find values for K and L which satisfy this equation. For example, dividing both sides by 10 and solving for K we get:

$$K = \frac{84}{10} - \frac{6L}{10}$$

$$K = 8.4 - 0.6L$$

Table 13.1 shows the combinations of capital and labour that satisfy this equation. For example, if 6 units of labour were used then $K = 8.4 - 0.6(6)$

$$K = 8.4 - 3.6$$

$$K = 4.8$$

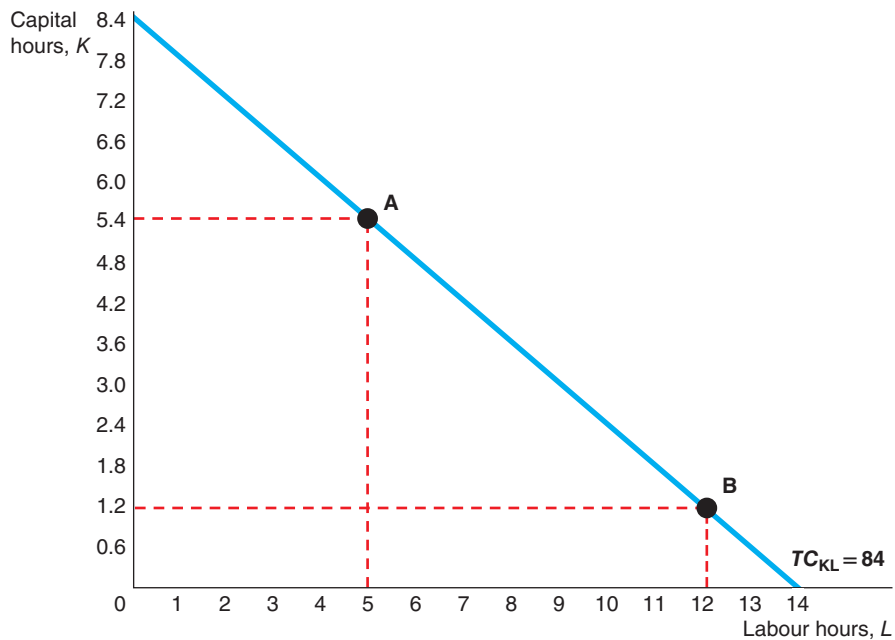
The information in Table 13.1 can be graphed as in Figure 13.3 with the number of capital hours on the vertical axis and the number of labour hours on the horizontal axis given the price of capital and labour at €10 per hour for capital and €6 per hour for labour. The isocost line $TC_{KL} = 84$ connects all the combinations of labour and capital to make pizzas which cost €84. At point A, 5.4 hours of capital and 5 hours of labour will have a total cost of €84 but so will the combination of 1.2 hours of capital and 12 hours of labour at point B.

TABLE 13.1**Factor Combinations to Satisfy the Equation $K = 8.4 - 0.5L$**

| K | L |
|-----|-----|
| 8.4 | 0 |
| 7.8 | 1 |
| 7.2 | 2 |
| 6.6 | 5 |
| 6.0 | 4 |
| 5.4 | 5 |
| 4.8 | 6 |
| 4.2 | 7 |
| 3.6 | 8 |
| 3.0 | 9 |
| 2.4 | 10 |
| 1.8 | 11 |
| 1.2 | 12 |
| 0.6 | 13 |
| 0 | 14 |

FIGURE 13.3**Isocost Lines**

Isocost lines connect combinations of capital and labour that a business can afford to buy given a budget constraint. The isocost line shown relates to a budget constraint of €84. With this budget constraint the factory owner could spend all the money on 8.4 hours of capital but would not be able to afford any workers giving the vertical intercept. If the business chose to spend the budget entirely on labour then it would be able to purchase 14 hours of labour per day but not be able to use any machines. Any point on the isocost line between these two extremes connects together combinations of capital and labour that could be purchased with the available budget. At point A, the factory owner could afford to buy 5.4 hours of capital and 5 hours of labour at the given price of capital and labour; at point B, she could afford to buy 1.2 hours of capital and 12 hours of labour.



Other isocost lines could be drawn connecting combinations of capital and labour at different levels of total cost. For each of these isocost lines, the vertical intercept shows how many units of capital the factory owner could buy with her budget constraint if she used zero hours of labour. The horizontal intercept shows how many hours of labour the factory owner could buy if zero hours of capital were purchased. The isocost line shows the combinations of capital and labour that could be purchased given the budget constraint.

The slope of the isocost line is the ratio of the price of labour to capital. As the isocost line is a straight line the slope is constant throughout. In this example, the slope is -0.6 . This tells us that in order not to increase her cost, for every one additional hour of labour employed she has to reduce the amount of capital by 0.6 hours and for every one additional hour of capital employed she must reduce labour by 1.66 hours.

SELF TEST What factors could cause the slope of the isoquant and isocost curves to change?

THE LEAST-COST INPUT COMBINATION

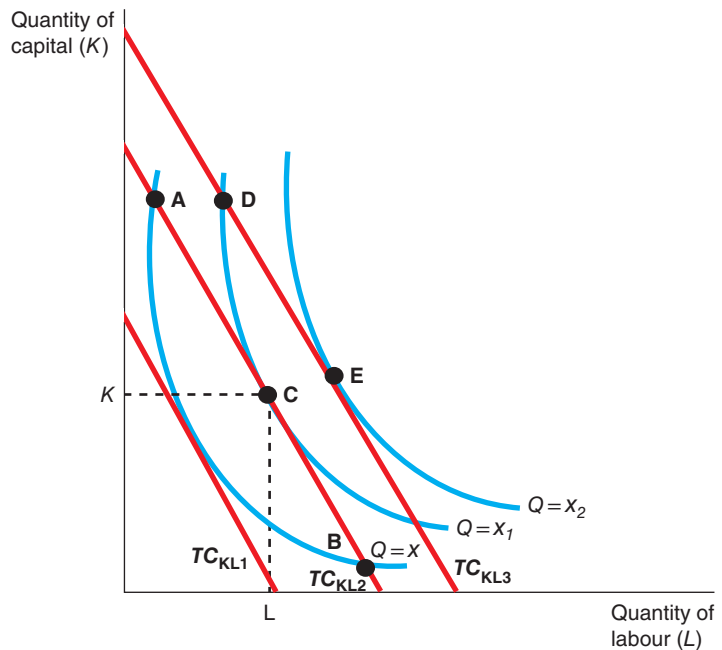
We now know the combination of factor inputs needed to produce given quantities of output (pizzas in our case) given by the isoquant curves and the cost of using different factor combinations given by the isocost lines. We can put these together to find the least cost input combination.

Figure 13.4 shows different isoquants relating to three different output levels $Q = x$, $Q = x_1$ and $Q = x_2$ and three isocost lines relating to three different cost levels for capital and labour giving the budget constraints TC_{KL1} , TC_{KL2} and TC_{KL3} .

FIGURE 13.4

The Least-Cost Input Combination

Three isoquants representing different output levels, $Q = x$, $Q = x_1$, and $Q = x_2$ are shown along with three different Isocost lines representing three different cost levels and budget constraints for the firm. The least-cost input combination is where the isoquant is tangential to the given budget constraint. Given a budget constraint of TC_{KL2} the least-cost combination of capital and labour is given as point C where the firm employs K hours of capital and L hours of labour.



Any point where the isocost line cuts the isoquant line is a possible combination of factors that could be used. The more resources a business has at its disposal the higher the output it can produce and this is determined by the amount of resources it can afford and hence by the prices of capital and labour. The question we want to ask is what the best or optimum combination is?

Let us assume that the factory owner has a budget constraint of TC_{KL2} . She could produce output $Q = x$ and employ the combination of factors of production at point A. Similarly she could use fewer hours of capital and more hours of labour and afford to produce the same output at point B. However, we could reasonably assume that if there was a way in which a business could use its existing budget and resources to produce more output then it would do so. It may make such a decision if it thought that it could sell more output.

Starting at point A, therefore, the factory owner could reduce the amount of capital used and increase the amount of labour to produce a higher output level $Q = x_1$ at point C. The owner cannot produce the output $Q = x_1$ using the combination of factors given by point D; however, because the owner does not have the funds to be able to afford this combination as it falls on a different isocost line, TC_{KL3} . She can afford to employ capital and labour in the combination given at point C. At point C the isoquant curve $Q = x_1$ is tangential to the isocost line TC_{KL2} . This is the least-cost input combination given the desired output level for the factory owner. At this point there is no incentive for her to change the combination of factors of production employed because to do so would mean that those resources would not be producing at maximum efficiency at minimum cost.

The owner might want to produce an output level given by $Q = x_2$ but given her budget constraint she cannot afford to produce that level of output. The optimum point, therefore, given existing productivity levels of factor inputs and the price of factor inputs, is C.

At this point of tangent, the point of least-cost input occurs where the marginal rate of technical substitution is equal to the ratio of the prices of factors. This is represented by the equation:

$$\frac{MP_L}{MP_K} = \frac{P_L}{P_K}$$

This is also sometimes written as:

$$\frac{MP_L}{P_K} = \frac{MP_K}{P_L}$$

SELF TEST Using Figure 13.4, explain what would happen if the price of labour hours fell but the price of capital hours remained the same. What would happen to the least-cost input combination?

Summary

Let us just summarize this section by thinking through this logically. If you were the factory owner faced with a budget constraint, you would want to ensure that you use your money in the best way possible to produce the maximum amount possible. Taking a factor input combination such as that at A, if there was a way in which you could reorganize those factor inputs so that they did not cost you any more but you could produce more pizzas, it would make sense to do so.

Cutting back on the use of capital and increasing labour means the additional output produced is greater but does not cost any more. Provided the benefit of doing this is greater than the cost incurred, it makes sense to make such a decision. If there is still a way to continue cutting capital use and increasing labour which would bring about increased production of pizzas, then it is clearly sensible to continue doing so until you reach a point where there is no benefit in shifting resources any further.

The least-cost input combination can change if the price of labour or capital changes (in which case the slope of the isocost line would change) or if both prices changed equally then the isocost line would shift either inwards or outwards depending on the direction of the price change. The shape of the isoquant curve might also change if the marginal productivity of either capital or labour changed.

Remember early in this analysis how we mentioned that this approach was a way of conceptualizing how businesses behave. The assumption is that firms want to maximize output at minimum cost. Firms will have some idea of the productivity of factor inputs and also of the cost of buying in factors. They will continually be looking to find ways to reorganize factors of production they employ to increase output but keep costs under control. The use of this model helps us to understand the logic behind business restructuring, outsourcing, seeking out cheaper suppliers, using different raw materials in different ways, spending money on training workers to be more effective in their jobs and other ways of influencing productivity, and helps explain why businesses are dynamic and constantly changing and evolving organizations.

CASE STUDY Labour Productivity

In the discussion of production in this chapter, productivity – output per unit of input per time period – is an important issue. In the UK in 2012, over half a million new jobs were created against the background of a stagnant economy. Labour inputs increased but if output is flat, this implies that the productivity of labour inputs is declining. Indeed, figures show that UK workers, output fell by 2.6 per cent per hour in the third quarter of 2012 compared to the first quarter in 2008. Whilst unemployment has not risen as far as commentators expected following the financial crisis, productivity has suffered with labour productivity 12.8 per cent below the pre-crisis level.

There have been various reasons put forward for this, including the fact that business investment is 16 per cent lower than the pre-crisis level so labour which is employed has fewer capital inputs to work with. The quality of labour has also been put forward as a factor in explaining the decline of labour productivity. The UK education system is falling behind in the provision of a quality education system, say critics. Standards of reading, science and maths have declined in comparison to other countries it is argued and as a result labour productivity is ultimately affected.

The amount and quality of capital goods is also a factor. Savings rates in the UK are among some of the lowest in the developed and developing world. There is a trade-off to be had: if UK citizens want an improved economy in the future, consumption has to be foregone in order to save more in order that capital investment can increase. With increasing levels of capital investment, the quality and quantity of capital can improve and as a result labour productivity will rise.



Levels of education have a significant impact on labour productivity. This is why politicians are concerned about improving standards of education and for the increasing focus on so-called STEM subjects (science, technology, engineering and maths).

CONCLUSION

In this chapter we have introduced a simplified technical analysis of how a firm might make production decisions based on it having two variable factors of production in the short run. The price of the factors of production and the budget of the firm determines the amount of factors that can be purchased and the factors can be used in different combinations to produce given amounts of output. Firms will constantly review the factor combinations they use in response to changing prices of factors and to changes in the productivity of factors to try and produce the maximum output at minimum cost. The least-cost input model allows us to conceptualize the changes that firms can make in response to changing factor market conditions.

IN THE NEWS

Production Decisions

Firms not only rely on making decisions themselves on production which can affect costs but also on the decisions of other firms. In this article we see how the production costs of small firms who manufacture goods for export are influenced by the decisions of firms involved in transport and distribution.

The Economics of Very Big Ships

Aboard one of the world's largest container ships, moving almost imperceptibly through the seas off Vietnam, it's easy to appreciate the economies of scale that allow a T-shirt made in China to be sent to the Netherlands for just 2.5 cents. The *Eleonora Maersk* and the other seven ships in her class are among the biggest ever built: almost 400m long, or the length of four football pitches, and another half-pitch across. The ship can carry 7,500 or so 40-foot containers, each of which can hold 70,000 T-shirts. On the voyage your correspondent took, the *Eleonora* was carrying Europe's New Year celebrations: 1,850 tonnes of fireworks, including 30 tonnes of gunpowder.

To move all this cargo from China to Europe in just over three weeks, the *Eleonora* boasts the largest internal-combustion engine ever built, as powerful as 1,000 family cars. This engine turns the longest propeller shaft (130m) ever made, at the end of which is the largest propeller, at 130 tonnes. Yet the ship is so automated that it requires a mere 13 people to crew it. Reassuringly, most captains prefer to take a few more.

Maersk Lines, the world's biggest container-shipping company and owner of the *Eleonora*, is betting that, given the current economics of world trade, the only way to go is yet

bigger. In February it announced an order for 20 even larger ships with a capacity of 18,000 twenty-foot-equivalent units (TEUs), the standard measure of container size. (The *Eleonora* can carry a mere 15,000.) The new ships will cost \$200m each. And, judging by this year's order books, Maersk's example will be followed by others. Singapore's Neptune Orient Lines has ordered ten vessels of 14,000 TEUs; Orient Overseas Container Line has ordered ten of 13,000 TEUs. This is excellent news for South Korean shipyards.

Most of these vessels will be designed for the Europe–Asia run – now the world's busiest trade route. Given the rising price of fuel, many shippers think they need huge ships to turn a profit. Fears of a

renewed slowdown in global trade accelerate the rush to find economies of scale.

Freight rates have plummeted in recent months, thanks to weakening demand and an oversupply of container ships, many of them ordered in the optimistic years before 2008. Freight rates on the Europe–Asia run are now below \$700 per TEU, less than half the peak a year ago. At this price, says Janet Lewis, an analyst at Macquarie, a bank, 'It's hard for any container company to be making any cash'.

Recent financial results bear this out. On 9 November 2011, Maersk said its container business had lost \$297m in the latest quarter. Neptune posted a third-quarter loss of \$91m, having made a profit of \$282m a year

Very large ships help reduce unit costs. Doubling the carrying capacity of a ship is unlikely to mean that costs will also double meaning the shipper can benefit from economies of scale.



(Continued)

earlier. Japan's 'K' Line reported a loss of \$239m on its container-shiping unit in the fiscal first half. And things could get a lot worse.

Companies are searching for new strategies to differentiate themselves from rivals, or merely to survive. Maersk has launched what it calls the 'Daily Maersk' service on the China–Europe run, deploying 70 vessels to promise daily deliveries to Felixstowe, Bremerhaven and Rotterdam, the three main European container ports. The firm is hoping to deliver 95 per cent of these containers on time, up from 80 per cent for its own service on that route. This would be far higher than the

industry average of about 65 per cent. If a container arrives more than a day late, Maersk pledges to compensate the customer. Orient Overseas, by contrast, is focusing on the quality of its cargo-handling. It charges more, for example, to ship perishable items, such as blood plasma.

Given the stormy waters that may well be ahead, it seems likely that shippers will seek economies of scale not only from bigger ships but also from mergers. The industry is too crowded, many analysts believe. Smaller firms may soon be swallowed like containers vanishing into the hold of the Eleonora Maersk.

Questions

- 1 How might cargo companies like Maersk use the principle of isoquants and isocosts to plan their production decisions?
- 2 Is a company such as Maersk likely to be labour or capital intensive? Explain.
- 3 Explain how economies of scale enable firms to transport products at very low unit cost.
- 4 Why do you think shipping firms are finding it difficult to be able to make profits?
- 5 How might differentiation help firms to be able to become more competitive.

SUMMARY

- The use of isoquants and isocosts helps to conceptualize the reasons why firms make decisions to change factor combinations used in production and how the prices of factor combinations can also influence those decisions.
- A firm's costs often depend on the time horizon being considered. In particular, many costs are fixed in the short run but variable in the long run. As a result, when the firm changes its level of production, average total cost may rise more in the short run than in the long run.

QUESTIONS FOR REVIEW

- 1 What are the constrained optimization problems facing firms?
- 2 Explain the difference between capital, labour and land intensive production and give an example of a firm in each case to illustrate the principle.
- 3 What is a production isoquant?
- 4 What is the marginal rate of technical substitution?
- 5 What is an isocost line?
- 6 What determines the slope of an isocost line?
- 7 Given the TC function $P_K K + P_L L = TC_{KL}$ where the price of capital is 50 and the price of labour is 10, what would the total cost of 12 units of capital and 8 units of labour be?
- 8 What is the least-cost input combination?
- 9 Define economies of scale and explain why they might arise.
- 10 If capital and labour were perfect substitutes, what would the shape of an isoquant be? Explain.

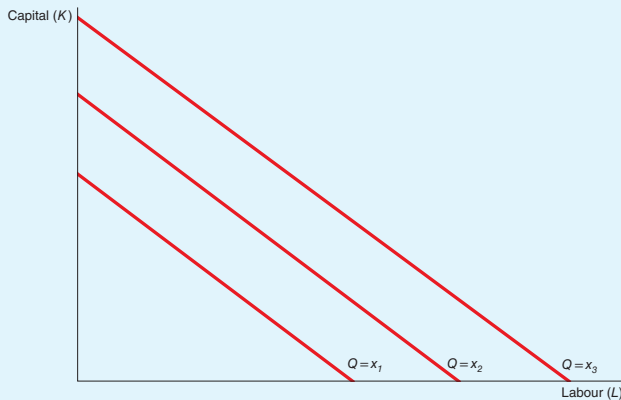
PROBLEMS AND APPLICATIONS

- 1 Isoquants are drawn as convex to the origin. Referring to the marginal rate of technical substitution, why do you think that isoquants are convex to the origin?
- 2 Look at the table below and sketch the isoquants implied by the data for the output levels $Q = 110$, $Q = 150$ and $Q = 180$.

| | | Labour (person hours per day) | | | | |
|---------------------------------|---|-------------------------------|-----|-----|-----|-----|
| | | 1 | 2 | 3 | 4 | 5 |
| Capital (machine hours per day) | 1 | 40 | 80 | 110 | 130 | 150 |
| | 2 | 80 | 120 | 150 | 170 | 210 |
| | 3 | 110 | 150 | 180 | 200 | 220 |
| | 4 | 130 | 170 | 200 | 220 | 230 |
| | 5 | 150 | 180 | 210 | 230 | 240 |

On your graph, put capital on the vertical axis and labour on the horizontal axis.

- 3 Using the graph you have constructed for Question 2, calculate the marginal rate of technical substitution for the output level $Q = 110$ when the firm moves from a combination of 3 machine hours of capital and 1 person hour of labour, to 1 machine hour of capital and 3 person hour of labour.
- 4 Look at the sketch of three production isoquants in the figure below. What do these isoquants tell you about the relationship between capital and labour in this particular instance?



- 5 Given the total cost function $50K + 12L = TC_{KL}$, calculate the total cost for a firm if it used the following combinations of capital and labour:
 - a. 5 units of capital and 8 units of labour
 - b. 10 units of labour and 3 units of capital
 - c. 7 units of capital and 12 units of labour.
- 6 Given the total cost function $50K + 12L = TC_{KL}$, at the following total cost levels, solve for K and find the factor combinations that satisfy the equation from $K = 1$ to $K = 5$.
 - a. $TC = 170$
 - b. $TC = 510$
 - c. $TC = 850$

- 7 If a firm faced the following situation:

$$\frac{MP_L}{TC_K} > \frac{MP_K}{TC_L}$$

What would be the incentives for the firm to change its production decisions? At what point would the firm stop changing its production decisions?

- 8 Sketch a diagram to show a firm's least-cost input combination. On your diagram show what would happen to the firm's optimum position if:
 - a. the price of capital increased but the price of labour stayed the same
 - b. the price of both labour and capital increased by the same amount
 - c. the price of labour and capital both increase but the price of capital increases by a greater amount than the price of labour.
- 9 A haulage firm uses containers to carry products for clients which currently measure 10m long by 2m wide by 4m high. The owner of the firm has decided to invest in new containers which double their length, width and height. Why might the owner have made this decision and what is likely to happen to: a. the total costs for the firm; and b. the average costs. Explain your reasoning.
- 10 Consider the following table of long-run total cost for three different firms:

| Quantity | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|-----|-----|-----|-----|------|------|------|
| Firm A | €60 | €70 | €80 | €90 | €100 | €110 | €120 |
| Firm B | 11 | 24 | 39 | 56 | 75 | 96 | 119 |
| Firm C | 21 | 34 | 49 | 66 | 85 | 106 | 129 |

Does each of these firms experience economies of scale or diseconomies of scale?

14

MARKET STRUCTURES I: MONOPOLY

IMPERFECT COMPETITION

So far in our analysis of firms we have assumed that conditions of perfect competition exist. What happens if these assumptions are dropped? If firms have some control over price, if products are not homogenous but are differentiated in some way, and if the market is dominated by firms who are able to exercise some element of market power, then firm behaviour will be different and the market outcomes will also be different to those we have analysed so far. When the assumptions of perfect competition do not hold, we say that firms are operating under imperfect competition. A firm operating under **imperfect competition** has the characteristic that it can differentiate its product in some way and so has some influence over the price it can charge for its product.

imperfect competition exists where firms are able to differentiate their product in some way and so can have some influence over price

There are different degrees of imperfect competition and we begin our analysis of firm behaviour under imperfect competition by looking at the opposite end of the competitive spectrum: monopoly. Strictly, a monopoly is a market structure with only one firm; however, in reality firms can exercise monopoly power by being the dominant firm in the market. Indeed, in many countries, firms might be investigated by regulators if they account for over 25 per cent of market share, with **market share** being the proportion of total sales in a market accounted for by a particular firm. The larger the market share a firm has the more market power it has. A firm can exercise **market power** when it is in a position to raise the price of its product and retain some sales – in other words, it is the opposite of the situation of a firm who is a *price taker*, it is a *price maker*. If a firm which is a price taker raises its price it loses all its sales as consumers will switch to the many rival firms that exist in the market (which are also producing identical products).

market share the proportion of total sales in a market accounted for by a particular firm
market power where a firm is able to raise the price of its product and not lose all its sales to rivals

Perhaps the most obvious example of how one firm can dominate a market is that of the Microsoft Corporation in the personal computer market. Microsoft produces its operating system, Windows, which accounts for around 82 per cent of the market (which now includes operating systems for desktop PCs, laptops and mobile devices). There are other firms who also provide operating systems such as Apple's iOS mobile device operating system, the open source Linux and Android, but Microsoft's Windows operating system dominates the market.

If a person or business wants to buy a copy of Windows, he or she has little choice but to give Microsoft the price that the firm has decided to charge for its product. Because of its ability to control the market for operating systems, Microsoft is said to have a monopoly in the market. In this situation the firm has few close competitors and, therefore, can influence the market price of its product.

In this chapter we examine the implications of this market power. We will see that market power alters the relationship between a firm's costs and the price at which it sells its product to the market. A competitive firm takes the price of its output as given by the market and then chooses the quantity it will supply so that price equals marginal cost. By contrast, the price charged by a monopoly exceeds marginal cost. This result is clearly true in the case of Microsoft's Windows. The marginal cost of Windows – the extra cost that Microsoft would incur by printing one more copy of the program onto a CD – is only a few euros. The market price of Windows is many times marginal cost.

It is perhaps not surprising that monopolies charge relatively high prices for their products. Customers of monopolies might seem to have little choice but to pay whatever the monopoly charges. But, if so, why is a copy of Windows priced at about €50 and not €500? Or €5,000? The reason, of course, is that if Microsoft set the price too high, fewer people would buy the product. People would buy fewer computers, switch to other operating systems or make illegal copies. Monopolies cannot achieve any level of profit they want because high prices reduce the amount that their customers buy. Although monopolies can control the prices of their goods, their profits are not unlimited.

As we examine the production and pricing decisions of monopolies, we also consider the implications of monopoly for society as a whole. Monopoly firms, like competitive firms, aim to maximize profit. But this goal has very different ramifications for competitive and monopoly firms. Under the assumptions of perfect competition, self-interested buyers and sellers in competitive markets are unwittingly led by an invisible hand to promote general economic well-being. By contrast, because monopoly firms are unchecked by competition, the outcome in a market with a monopoly is often not in the best interest of society.

As we examine the problems that monopolies raise for society, we will also discuss the various ways in which government policymakers might respond to these problems.

WHY MONOPOLIES ARISE

A firm is a **monopoly** if it is the sole seller of its product and if its product does not have close substitutes. Whilst this is the strict definition of a monopoly, as we have seen firms are said to have monopoly power if they are a dominant seller in the market and are able to exert some control over the market as a result. In the analysis that follows, however, the assumption is that there is only one seller.

monopoly a firm that is the sole seller of a product without close substitutes

The fundamental cause of monopoly is **barriers to entry**: a barrier to entry exists where there is something that prevents a firm from entering an industry. The stronger the barriers to entry the more difficult it is for a firm to enter a market and the more market power a firm can exert. Remember, when we analysed firm behaviour in perfect competition we assumed there was free entry and exit to the market.

barriers to entry anything which prevents a firm from entering a market or industry

A monopoly can remain the only seller in its market because other firms cannot enter the market and compete with it. Barriers to entry, in turn, have four main sources:

- A key resource is owned by a single firm.
- The government gives a single firm the exclusive right to produce some good or service.
- The costs of production make a single producer more efficient than a large number of producers.
- A firm is able to gain control of other firms in the market and thus grow in size.

Let's briefly discuss each of these.

Monopoly Resources

The simplest way for a monopoly to arise is for a single firm to own a key resource. For example, consider the market for water in a small town on a remote island not served by the water company from the mainland. If dozens of town residents on the island have working wells, the competitive model discussed in the preceding chapter describes the behaviour of sellers. As a result, the price of a litre of water is driven to equal the marginal cost of pumping an extra litre. But if there is only one well in town and it is impossible to get water from anywhere else, then the owner of the well has a monopoly on water. Not surprisingly, the monopolist has much greater market power than any single firm in a competitive market. In the case of a necessity like water, the monopolist could command quite a high price, even if the marginal cost is low.

Although exclusive ownership of a key resource is a potential cause of monopoly, in practice monopolies rarely arise for this reason. Actual economies are large, and resources are owned by many people. Indeed, because many goods are traded internationally, the natural scope of their markets is often worldwide. There are, therefore, few examples of firms that own a resource for which there are no close substitutes.

Government-Created Monopolies

In many cases, monopolies arise because the government has given one person or firm the exclusive right to sell some good or service. Sometimes the monopoly arises from the sheer political clout of the would-be monopolist. European kings, for example, once granted exclusive business licences to their friends and allies in order to raise money – a highly prized monopoly being the exclusive right to sell and distribute salt in a particular region of Europe. Even today, governments sometimes grant a monopoly (perhaps even to themselves) because doing so is viewed to be in the public interest. In Sweden, for example, the retailing of alcoholic beverages is carried out under a state-owned monopoly known as the Systembolaget, because the Swedish government deems it to be in the interests of public health to be able to control directly the sale of alcohol. As a member of the EU, questions have been raised about this policy but Sweden seems keen to maintain its control of alcohol sales.

In a study commissioned by the Swedish National Institute for Public Health in 2007, researchers concluded that if retail alcohol sales were privatized, the net effects on the country would be negative with an increase in alcohol-related illness and deaths, fatal accidents, suicides and homicides and a large increase in the number of working days lost to sickness. (See Holder, H. (Ed.) (2007) *If Retail Alcohol Sales in Sweden were Privatized, What Would be the Potential Consequences?* http://www.systembolagetkampanj.se/forskarrapport_en/downloads/Hela_rapporten.pdf)

The patent and copyright laws are two important examples of how the government creates a monopoly to serve the public interest. When a pharmaceutical company discovers a new drug, it can apply to the government for a patent. If the government deems the drug to be truly original, it approves the **patent**, which gives the company the exclusive right to manufacture and sell the drug for a fixed number of years – often 20 years. A patent, therefore, is a means of establishing and enforcing property rights.

patent the right conferred on the owner to prevent anyone else making or using an invention or manufacturing process without permission

Similarly, when a novelist finishes a book, he can **copyright** it. The copyright is a government guarantee that no one can print and sell the work without the author's permission and thus conveys rights to the owner to control how his work is used. Copyright relates to the expression of ideas where some judgement or skill is used to create the work. It covers the creation of literary works, music, the arts, sound recordings, broadcasts, films and typographical arrangements of published material. The copyright makes the novelist a monopolist in the sale of his novel and is a means of establishing intellectual property rights.

copyright the right of an individual or organization to own things they create in the same way as a physical object to prevent others from copying or reproducing the creation

The effects of patent and copyright laws are easy to see. Because these laws give one producer a monopoly, they lead to higher prices than would occur under competition. But by allowing these monopoly producers to charge higher prices and earn higher profits, the laws also encourage some desirable behaviour. Drug companies are allowed to be monopolists in the drugs they discover in order to encourage research. Authors are allowed to be monopolists in the sale of their books to encourage them to write more and better books.

Thus, the laws governing patents and copyrights have benefits and costs. The benefits of the patent and copyright laws are the increased incentive for creative activity. These benefits are offset, to some extent, by the costs of monopoly pricing, which we examine fully later in this chapter.

Natural Monopolies

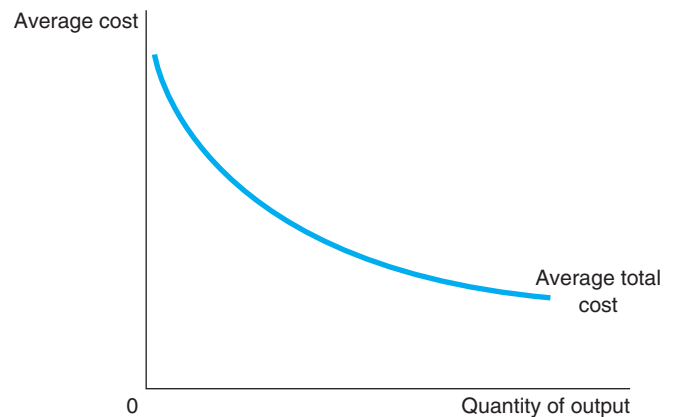
An industry is a **natural monopoly** when a single firm can supply a good or service to an entire market at a lower cost than could two or more firms. A natural monopoly arises when there are economies of scale over the relevant range of output. Figure 14.1 shows the average total costs of a firm with economies of scale. In this case, a single firm can produce any amount of output at least cost. That is, for any given amount of output, a larger number of firms leads to less output per firm and higher average total cost.

natural monopoly a monopoly that arises because a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms

FIGURE 14.1

Economies of Scale as a Cause of Monopoly

When a firm's average total cost curve continually declines, the firm has what is called a natural monopoly. In this case, when production is divided among more firms, each firm produces less, and average total cost rises. As a result, a single firm can produce any given amount at the smallest cost.



An example of a natural monopoly is the distribution of water. To provide water to residents of a town, a firm must build a network of pipes throughout the town. If two or more firms were to compete in the provision of this service, each firm would have to pay the fixed cost of building a network. Thus, the average total cost of water is lowest if a single firm serves the entire market.

Some goods in the economy are excludable but not rival as noted in Chapter 10. An example is a bridge used so infrequently that it is never congested. The bridge is excludable because a toll collector can prevent someone from using it. The bridge is not rival because use of the bridge by one person does not diminish the ability of others to use it. Because there is a fixed cost of building the bridge and a negligible

marginal cost of additional users, the average total cost of a trip across the bridge (the total cost divided by the number of trips) falls as the number of trips rises. Hence, the bridge is a natural monopoly.

When a firm is a natural monopoly, it is less concerned about new entrants eroding its monopoly power. Normally, a firm has trouble maintaining a monopoly position without ownership of a key resource or protection from the government. The monopolist's profit attracts entrants into the market, and these entrants make the market more competitive. By contrast, entering a market in which another firm has a natural monopoly is unattractive. Would-be entrants know that they cannot achieve the same low costs that the monopolist enjoys because, after entry, each firm would have a smaller piece of the market.

In some cases, the size of the market is one determinant of whether an industry is a natural monopoly. Again, consider a bridge across a river. When the population is small, the bridge may be a natural monopoly. A single bridge can satisfy the entire demand for trips across the river at lowest cost. Yet as the population grows and the bridge becomes congested, satisfying the entire demand may require two or more bridges across the same river. Thus, as a market expands, a natural monopoly can evolve into a competitive market.

External Growth

Many of the largest firms in the world have grown partly through acquisition, merger or takeover of other firms. As they do so, the industry becomes more concentrated; there are fewer firms in the industry. One effect of this is that a firm might be able to develop monopoly power over its rivals and erect barriers to entry to make it harder for new firms to enter. It is for this reason that governments monitor such acquisitions to see if there are implications for competition.

SELF TEST What are the four reasons that a market might have a monopoly? Give three examples of monopolies, and explain the reason for each.

HOW MONOPOLIES MAKE PRODUCTION AND PRICING DECISIONS

Now that we know how monopolies arise, we can consider how a monopoly firm decides how much of its product to make and what price to charge for it. The analysis of monopoly behaviour in this section is the starting point for evaluating whether monopolies are desirable and what policies the government might pursue in monopoly markets.

Monopoly versus Competition

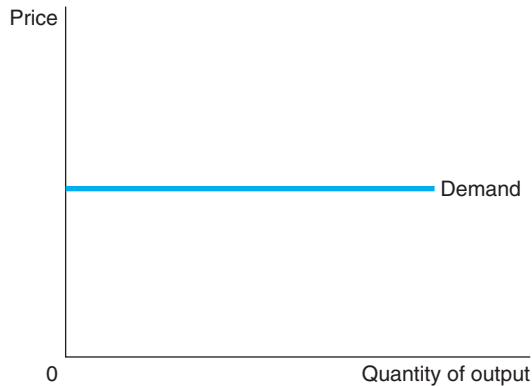
The key difference between a competitive firm and a monopoly is the monopoly's ability to influence the price of its output. A competitive firm is small relative to the market in which it operates and, therefore, takes the price of its output as given by market conditions. By contrast, because a monopoly is the sole producer in its market, it can alter the price of its good by adjusting the quantity it supplies to the market.

One way to view this difference between a competitive firm and a monopoly is to consider the demand curve that each firm faces. When we analysed profit maximization by competitive firms, we drew the market price as a horizontal line. Because a competitive firm can sell as much or as little as it wants at this price, the competitive firm faces a horizontal demand curve, as in panel (a) of Figure 14.2. If it charged any price above P_0 , however, it would lose all of its sales to its rivals. In effect, because the competitive firm sells a product with many perfect substitutes (the identical products of all the other firms in its market), the demand curve that any one firm faces is perfectly elastic. The firm has no market power.

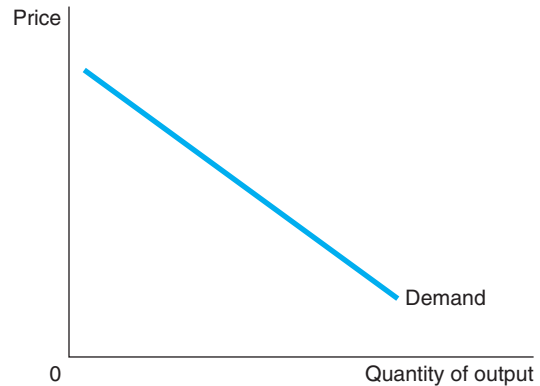
By contrast, because a monopoly is the sole producer in its market, its demand curve is the market demand curve. Thus, the monopolist's demand curve slopes downwards for all the usual reasons, as in

FIGURE 14.2**Demand Curves for Competitive and Monopoly Firms**

Because competitive firms are price takers, they in effect face horizontal demand curves, as in panel (a). In such a situation the firm is a price taker and has no market power. Because a monopoly firm is the sole producer in its market, it faces the downward sloping market demand curve, as in panel (b). As a result, the monopoly has to accept a lower price if it wants to sell more output.



(a) A competitive firm's demand curve



(b) A monopolist's demand curve

panel (b) of Figure 14.2. If the monopolist raises the price of its good, consumers buy less of it. Looked at another way, if the monopolist reduces the quantity of output it sells, the price of its output increases.

The market demand curve provides a constraint on a monopoly's ability to profit from its market power. A monopolist would prefer, if it were possible, to charge a high price and sell a large quantity at that high price. The market demand curve makes that outcome impossible. In particular, the market demand curve describes the combinations of price and quantity that are available to a monopoly firm. By adjusting the quantity produced (or, equivalently, the price charged), the monopolist can choose any point on the demand curve, but it cannot choose a point off the demand curve.

What point on the demand curve will the monopolist choose? As with competitive firms, we assume that the monopolist's goal is to maximize profit. Because the firm's profit is total revenue minus total costs, our next task in explaining monopoly behaviour is to examine a monopolist's revenue.

A Monopoly's Revenue

Consider a town with a single producer of water. Table 14.1 shows how the monopoly's revenue might depend on the amount of water produced.

The first two columns show the monopolist's demand schedule. If the monopolist produces just 1 litre of water, it can sell that litre for €1. If it produces 2 litres, it must lower the price to €0.90 in order to sell both litres. And if it produces 3 litres, it must lower the price to €0.80 and so on. If you graphed these two columns of numbers, you would get a typical downwards sloping demand curve.

The third column of the table presents the monopolist's *total revenue*. It equals the quantity sold (from the first column) times the price (from the second column). The fourth column computes the firm's *average revenue*, the amount of revenue the firm receives per unit sold. We compute average revenue by taking the number for total revenue in the third column and dividing it by the quantity of output in the first column. As we discussed in the previous chapter, average revenue always equals the price of the good. This is true for monopolists as well as for competitive firms.

The last column of Table 14.1 computes the firm's *marginal revenue*, the amount of revenue that the firm receives for each additional unit of output. We compute marginal revenue by taking the change in total revenue when output increases by 1 unit. For example, when the firm is producing 3 litres of water it receives total revenue of €2.40. Raising production to 4 litres increases total revenue to €2.80. Thus, marginal revenue is €2.80 minus €2.40 or €0.40.

TABLE 14.1 A Monopoly's Total, Average and Marginal Revenue

| Quantity of water (Q) | Price (P) | Total revenue ($TR = P \times Q$) | Average revenue ($AR = TR/Q$) | Marginal revenue ($MR = \Delta TR/\Delta Q$) |
|---------------------------|---------------|-------------------------------------|---------------------------------|--|
| 0 litres | €1.1 | €0 | – | €1 |
| 1 | 1.0 | 1.0 | €1 | 0.8 |
| 2 | 0.9 | 1.8 | 0.9 | 0.6 |
| 3 | 0.8 | 2.4 | 0.8 | 0.4 |
| 4 | 0.7 | 2.8 | 0.7 | 0.2 |
| 5 | 0.6 | 3.0 | 0.6 | 0 |
| 6 | 0.5 | 3.0 | 0.5 | -0.2 |
| 7 | 0.4 | 2.8 | 0.4 | -0.4 |
| 8 | 0.3 | 2.4 | 0.3 | -0.4 |

Table 14.1 shows a result that is important for understanding monopoly behaviour: a monopolist's marginal revenue is always less than the price of its good. For example, if the firm raises production of water from 3 to 4 litres, it will increase total revenue by only €0.40, even though it will be able to sell each litre for €0.70. For a monopoly, marginal revenue is lower than price because a monopoly faces a downwards sloping demand curve. To increase the amount sold, a monopoly firm must lower the price of its good. Hence, to sell the fourth litre of water, the monopolist must get less revenue for each of the first 3 litres.

Marginal revenue for monopolies is very different from marginal revenue for competitive firms. When a monopoly increases the amount it sells, it has two effects on total revenue ($P \times Q$):

- *The output effect.* More output is sold, so Q is higher, which tends to increase total revenue.
- *The price effect.* The price falls, so P is lower, which tends to decrease total revenue.

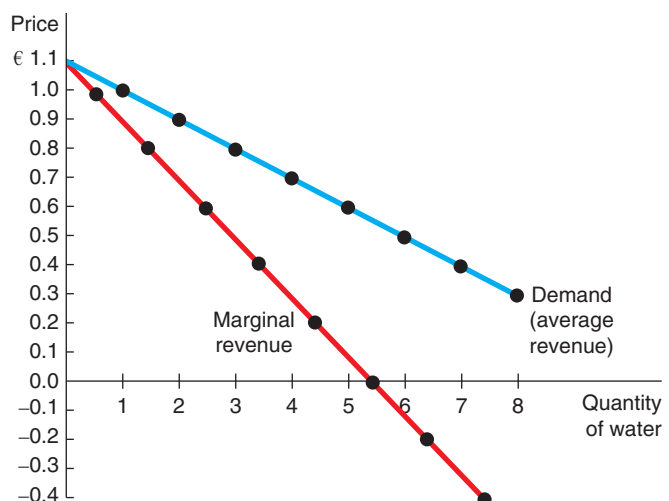
Because a competitive firm can sell all it wants at the market price, there is no price effect. When it increases production by 1 unit, it receives the market price for that unit, and it does not receive any less for the units it was already selling. That is, because the competitive firm is a price taker, its marginal revenue equals the price of its good. By contrast, when a monopoly increases production by 1 unit, it must reduce the price it charges for every unit it sells, and this cut in price reduces revenue on the units it was already selling. As a result, a monopoly's marginal revenue is less than its price.

Figure 14.3 graphs the demand curve and the marginal revenue curve for a monopoly firm. (Because the firm's price equals its average revenue, the demand curve is also the average-revenue curve.) These two curves always start at the same point on the vertical axis because the marginal revenue of the first unit sold equals the price of the good. But thereafter, for the reason we just discussed, the monopolist's marginal revenue is less than the price of the good. Thus, a monopoly's marginal revenue curve lies below its demand curve.

FIGURE 14.3

Demand and Marginal Revenue Curves for a Monopoly

The demand curve shows how the quantity affects the price of the good. The marginal revenue curve shows how the firm's revenue changes when the quantity increases by 1 unit. Because the price on all units sold must fall if the monopoly increases production, marginal revenue is always less than the price.



You can see in the figure (as well as in Table 14.1) that marginal revenue can even become negative. Marginal revenue is negative when the price effect on revenue is greater than the output effect. In this case, when the firm produces an extra unit of output, the price falls by enough to cause the firm's total revenue to decline, even though the firm is selling more units.

Profit Maximization

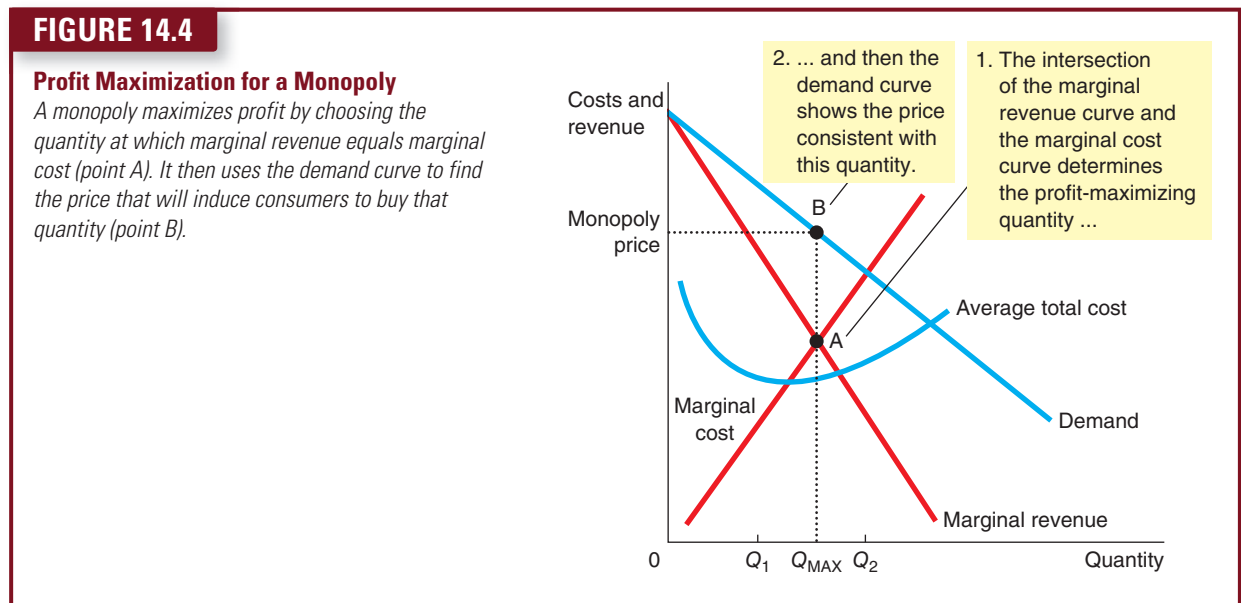
Now that we have considered the revenue of a monopoly firm, we are ready to examine how such a firm maximizes profit.

Figure 14.4 graphs the demand curve, the marginal revenue curve and the cost curves for a monopoly firm. These curves contain all the information we need to determine the level of output that a profit-maximizing monopolist will choose.

Suppose, first, that the firm is producing at a low level of output, such as Q_1 . In this case, marginal cost is less than marginal revenue. If the firm increased production by 1 unit, the additional revenue would exceed the additional costs, and profit would rise. Thus, when marginal cost is less than marginal revenue, the firm can increase profit by producing more units.

A similar argument applies at high levels of output, such as Q_2 . In this case, marginal cost is greater than marginal revenue. If the firm reduced production by 1 unit, the costs saved would exceed the revenue lost. Thus, if marginal cost is greater than marginal revenue, the firm can raise profit by reducing production.

In the end, the firm adjusts its level of production until the quantity reaches Q_{MAX} , at which marginal revenue equals marginal cost. Thus, the monopolist's profit-maximizing quantity of output is determined by the intersection of the marginal revenue curve and the marginal cost curve. In Figure 14.4, this intersection occurs at point A.



Remember that competitive firms maximize profit at the quantity of output at which marginal revenue equals marginal cost. In following this rule for profit maximization, competitive firms and monopolies are alike. But there is also an important difference between these types of firm: the marginal revenue of a competitive firm equals its price, whereas the marginal revenue of a monopoly is less than its price. That is:

For a competitive firm: $P = MR = MC$

For a monopoly firm: $P > MR = MC$

FYI



Why a Monopoly Does Not Have a Supply Curve

You may have noticed that we have analysed the price in a monopoly market using the market demand curve and the firm's cost curves. We have not made any mention of the market supply curve.

What happened to the supply curve? Although monopoly firms make decisions about what quantity to supply (in the way described in this chapter), a monopoly does not have a supply curve. A supply curve tells us

the quantity that firms choose to supply at any given price. This concept makes sense when we are analysing competitive firms, which are price takers. But a monopoly firm is a price maker, not a price taker. It is not meaningful to ask what such a firm would produce at any price because the firm sets the price at the same time it chooses the quantity to supply.

Indeed, the monopolist's decision about how much to supply is

impossible to separate from the demand curve it faces. The shape of the demand curve determines the shape of the marginal revenue curve, which in turn determines the monopolist's profit-maximizing quantity. In a competitive market, supply decisions can be analysed without knowing the demand curve, but that is not true in a monopoly market. Therefore, we never talk about a monopoly's supply curve.

The equality of marginal revenue and marginal cost at the profit-maximizing quantity is the same for both types of firm. What differs is the relationship of the price to marginal revenue and marginal cost.

How does the monopoly find the profit-maximizing price for its product? The demand curve answers this question because the demand curve relates the amount that customers are willing to pay to the quantity sold. Thus, after the monopoly firm chooses the quantity of output that equates marginal revenue and marginal cost, it uses the demand curve to find the price consistent with that quantity. In Figure 14.4, the profit-maximizing price is found at point B.

We can now see a key difference between markets with competitive firms and markets with a monopoly firm: in competitive markets, price equals marginal cost. In monopolized markets, price exceeds marginal cost. As we will see in a moment, this finding is crucial to understanding the social cost of monopoly.

A Monopoly's Profit

How much profit does the monopoly make? To see the monopoly's profit, recall that profit (π) equals total revenue (TR) minus total costs (TC):

$$\pi = TR - TC$$

We can rewrite this as:

$$\pi = (TR/Q - TC/Q) \times Q$$

TR/Q is average revenue, which equals the price P and TC/Q is average total cost ATC . Therefore:

$$\pi = (P - ATC) \times Q$$

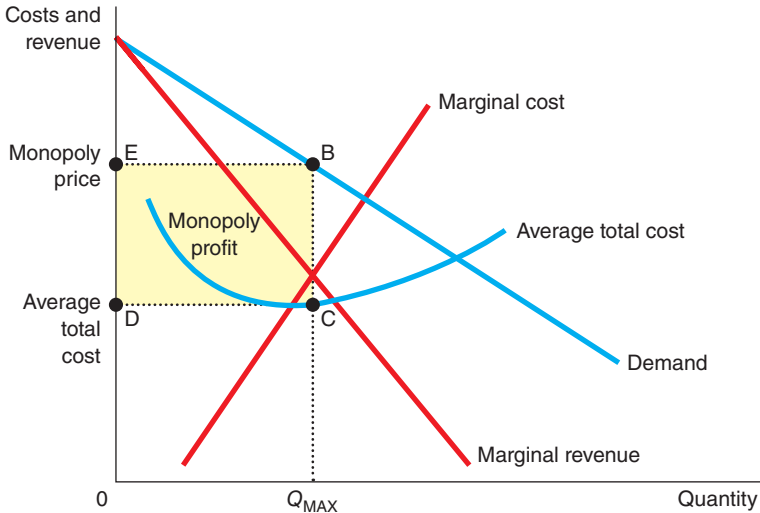
This equation for profit (which is the same as the profit equation for competitive firms) allows us to measure the monopolist's profit in our graph.

Consider the shaded box in Figure 14.5. The height of the box (the segment BC) is price minus average total cost, $P - ATC$, which is the profit on the typical unit sold. The width of the box (the segment DC) is the quantity sold Q_{MAX} . Therefore, the area of this box is the monopoly firm's total profit.

FIGURE 14.5

The Monopolist's Profit

The area of the box BCDE equals the profit of the monopoly firm. The height of the box (BC) is price minus average total cost, which equals profit per unit sold. The width of the box (DC) is the number of units sold.



CASE STUDY

Monopoly Drugs versus Generic Drugs

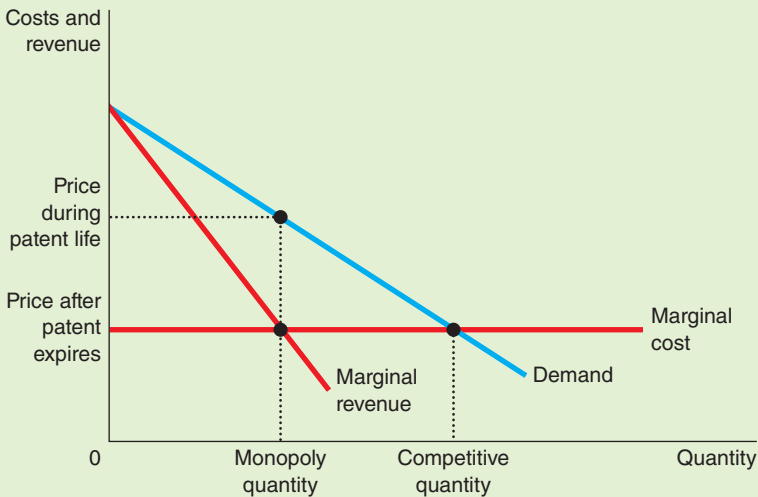
According to our analysis, prices are determined quite differently in markets where a firm has a monopoly compared to the way they are in competitive markets. A natural place to test this theory is the market for pharmaceutical drugs, because this market takes on both market structures. When a firm discovers a new drug, patent laws give the firm a monopoly on the sale of that drug. But eventually the firm's patent runs out, and any company can make and sell the drug. At that time, the market switches from being a monopoly to being competitive.

What should happen to the price of a drug when the patent runs out? Figure 14.6 shows the market for a typical drug. In this figure, the marginal cost of producing the drug is constant. (This is approximately true for many drugs.)

FIGURE 14.6

The Market for Drugs

When a patent gives a firm a monopoly over the sale of a drug, the firm charges the monopoly price, which is well above the marginal cost of making the drug. When the patent on a drug runs out, new firms enter the market, making it more competitive. As a result, the price falls from the monopoly price to marginal cost.



(Continued)

During the life of the patent, the monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost and charging a price well above marginal cost. But when the patent runs out, the profit from making the drug should encourage new firms to enter the market. As the market becomes more competitive, the price should fall to equal marginal cost.

Experience is, in fact, consistent with our theory. When the patent on a drug expires, other companies quickly enter and begin selling so-called generic products that are chemically identical to the former monopolist's brand-name product. And just as our analysis predicts, the price of the competitively produced generic drug is well below the price that the monopolist was charging.

The expiration of a patent, however, does not cause the monopolist to lose all its market power. Some consumers remain loyal to the brand-name drug, perhaps out of fear that the new generic drugs are not actually the same as the drug they have been using for years. As a result, the former monopolist can continue to charge a price somewhat above the price charged by its new competitors.



Pfizer held the patent for the drug Viagra until July 2013. When the patent ran out, the market for this particular drug could be flooded with generic versions priced much lower than Pfizer's Viagra.

SELF TEST Can a monopolist determine both the price it chooses to charge and the amount it sells? Why? Why not?

THE WELFARE COST OF MONOPOLY

Is monopoly a good way to organize a market? We have seen that a monopoly, in contrast to a competitive firm, charges a price above marginal cost. From the standpoint of consumers, this high price makes monopoly undesirable. At the same time, however, the monopoly is earning profit from charging this high price. From the standpoint of the owners of the firm, the high price makes monopoly very desirable. Is it possible that the benefits to the firm's owners exceed the costs imposed on consumers, making monopoly desirable from the standpoint of society as a whole?

We can answer this question using total surplus as our measure of economic well-being. Recall that total surplus is the sum of consumer surplus and producer surplus. In this case, there is a single producer – the monopolist.

You might already be able to guess the result of this analysis. We have noted that the equilibrium of supply and demand in a competitive market is not only a natural outcome but a desirable one. The invisible hand of the market leads to an allocation of resources that makes total surplus as large as it can be. Because a monopoly leads to an allocation of resources different from that in a competitive market, the outcome must, in some way, fail to maximize total economic well-being.

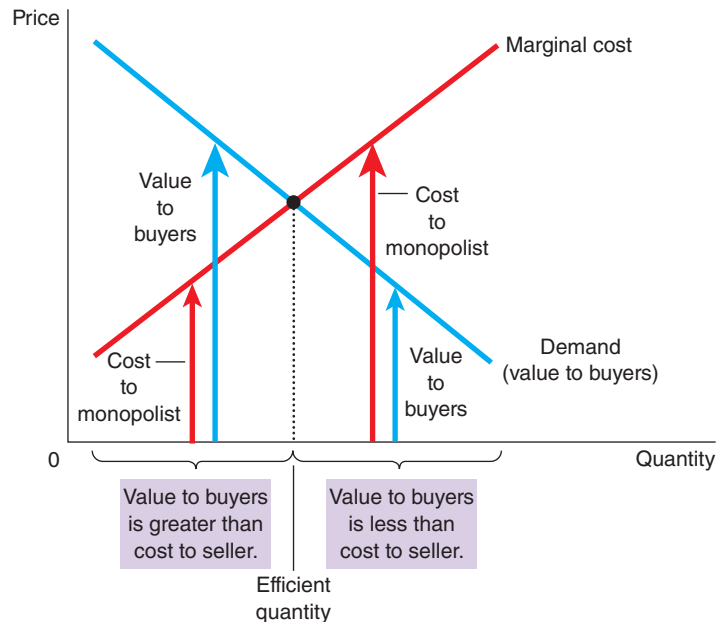
The Deadweight Loss

The fact that the market outcome under monopoly is different to that under conditions of perfect competition means there is a deadweight loss associated with monopoly. Total surplus equals the value of the good to consumers minus the costs of making the good incurred by the monopoly producer. In Figure 14.7 the demand curve reflects the value of the good to consumers, as measured by their willingness to pay for it. The marginal cost curve reflects the costs of the monopolist. Thus, the socially efficient quantity is found where the demand curve and the marginal cost curve intersect. Below this quantity, the value to consumers exceeds the marginal cost of providing the good, so increasing output would raise total surplus. Above this quantity, the marginal cost exceeds the value to consumers, so decreasing output would raise total surplus.

FIGURE 14.7

The Efficient Level of Output

Total surplus in the market would be maximized at the level of output where the demand curve and marginal cost curve intersect. Below this level, the value of the good to the marginal buyer (as reflected in the demand curve) exceeds the marginal cost of making the good. Above this level, the value to the marginal buyer is less than marginal cost.



The efficient outcome would be where the demand curve intersects the marginal cost curve where $P = MC$. Because this price would give consumers an accurate signal about the cost of producing the good, consumers would buy the efficient quantity. We can evaluate the welfare effects of monopoly by comparing the level of output that the monopolist chooses to the socially efficient output where $P = MC$.

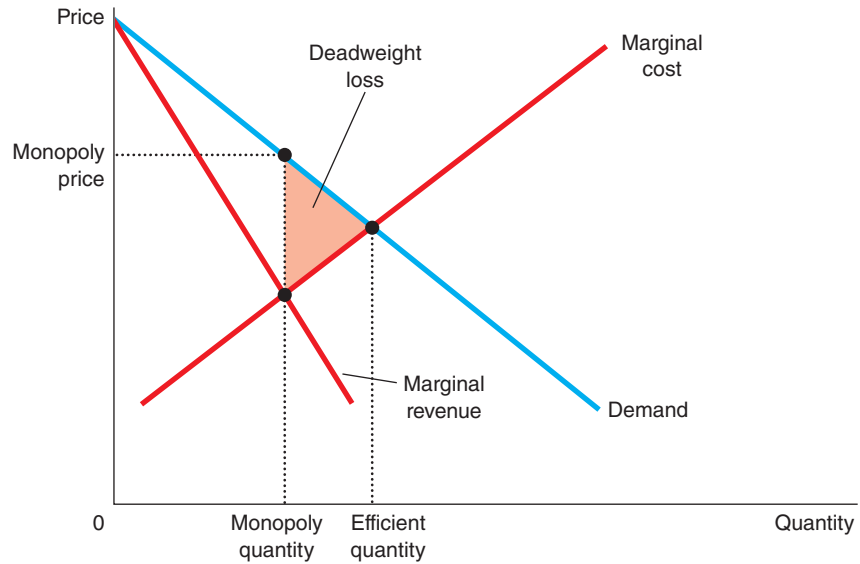
The monopolist chooses the profit maximizing output where the marginal revenue and marginal cost curves intersect but this is not the same as the socially efficient output where the demand and marginal cost curves intersect. Figure 14.8 shows the comparison. The monopolist produces less than the socially efficient quantity of output.

We can also view the inefficiency of monopoly in terms of the monopolist's price. Because the market demand curve describes a negative relationship between the price and quantity of the good, a quantity that is inefficiently low is equivalent to a price that is inefficiently high. When a monopolist charges a price above marginal cost, some potential consumers value the good at more than its marginal cost but less than the monopolist's price. These consumers do not end up buying the good. Because the value these consumers place on the good is greater than the cost of providing it to them, this result is inefficient. Thus, monopoly pricing prevents some mutually beneficial trades from taking place.

Figure 14.8 shows the deadweight loss. Recall that the demand curve reflects the value to consumers and the marginal cost curve reflects the costs to the monopoly producer. Thus, the area of the deadweight loss triangle between the demand curve and the marginal cost curve equals the total surplus lost because of monopoly pricing.

FIGURE 14.8**The Inefficiency of Monopoly**

Because a monopoly charges a price above marginal cost, not all consumers who value the good at more than its cost buy it. Thus, the quantity produced and sold by a monopoly is below the socially efficient level. The deadweight loss is represented by the area of the triangle between the demand curve (which reflects the value of the good to consumers) and the marginal cost curve (which reflects the costs of the monopoly producer).



The deadweight loss caused by monopoly is similar to the deadweight loss caused by a tax. Indeed, a monopolist is like a private tax collector. A tax on a good, remember, places a wedge between consumers' willingness to pay (as reflected in the demand curve) and producers' costs (as reflected in the supply curve). Because a monopoly exerts its market power by charging a price above marginal cost, it places a similar wedge. In both cases, the wedge causes the quantity sold to fall short of the social optimum. The difference between the two cases is that the government gets the revenue from a tax, whereas a private firm gets the monopoly profit.

The Monopoly's Profit: A Social Cost?

It is tempting to decry monopolies for 'profiteering' at the expense of the public. And, indeed, a monopoly firm does earn a higher profit by virtue of its market power. According to the economic analysis of monopoly, however, the firm's profit is not in itself necessarily a problem for society.

Welfare in a monopolized market, like all markets, includes the welfare of both consumers and producers. Whenever a consumer pays an extra euro to a producer because of a monopoly price, the consumer is worse off by a euro, and the producer is better off by the same amount. This transfer from the consumers of the good to the owners of the monopoly does not affect the market's total surplus – the sum of consumer and producer surplus. In other words, the monopoly profit itself does not represent a shrinkage in the size of the economic pie; it merely represents a bigger slice for producers and a smaller slice for consumers. Unless consumers are for some reason more deserving than producers – a judgement that goes beyond the realm of economic efficiency – the monopoly profit is not a social problem.

The problem in a monopolized market arises because the firm produces and sells a quantity of output below the level that maximizes total surplus. The deadweight loss measures how much the economic pie shrinks as a result. This inefficiency is connected to the monopoly's high price: consumers buy fewer units when the firm raises its price above marginal cost. But keep in mind that the profit earned on the units that continue to be sold is not the problem. The problem stems from the inefficiently low quantity of output. Put differently, if the high monopoly price did not discourage some consumers from buying the good, it would raise producer surplus by exactly the amount it reduced consumer surplus, leaving total surplus unchanged from the socially efficient outcome.

There is, however, a possible exception to this conclusion. Suppose that a monopoly firm has to incur additional costs to maintain its monopoly position. For example, a firm with a government-created monopoly might need to hire lobbyists to convince lawmakers to continue its monopoly. In this case, the

monopoly may use up some of its monopoly profits paying for these additional costs. If so, the social loss from monopoly includes both these costs and the deadweight loss resulting from a price above marginal cost.

SELF TEST Without the conferring of monopoly power through barriers to entry such as patents, would any competitive firm produce goods which have wider social benefits such as pharmaceutical drugs? Does this sort of consideration have to be taken into account when looking at the welfare loss associated with monopolies?

PRICE DISCRIMINATION

So far we have been assuming that the monopoly firm charges the same price to all customers. Yet in many cases firms try to sell the same good to different customers for different prices, even though the costs of producing for the two customers are the same. This practice is called **price discrimination**.

price discrimination the business practice of selling the same good at different prices to different customers

Before discussing the behaviour of a price-discriminating monopolist, we should note that price discrimination is not possible when a good is sold in a competitive market. In a competitive market, there are many firms selling the same good at the market price. No firm is willing to charge a lower price to any customer because the firm can sell all it wants at the market price. And if any firm tried to charge a higher price to a customer, that customer would buy from another firm. For a firm to price discriminate, it must have some market power.

A Parable about Pricing

To understand why a monopolist would want to price discriminate, let's consider a simple example. Imagine that you are the chief executive officer (CEO) of Readalot Publishing Company. Readalot's best-selling author has just written her latest novel. To keep things simple, let's imagine that you pay the author a flat €2 million for the exclusive rights to publish the book. Let's also assume – for simplicity – that the cost of printing the book is zero. Readalot's profit, therefore, is the revenue it gets from selling the book minus the €2 million it has paid to the author. Given these assumptions, how would you, as Readalot's CEO, decide what price to charge for the book?

Your first step in setting the price is to estimate what the demand for the book is likely to be. Readalot's marketing department tells you that the book will attract two types of readers. The book will appeal to the author's 100,000 die-hard fans. These fans will be willing to pay as much as €30 for the book. In addition, the book will appeal to about 400,000 less enthusiastic readers who will be willing to pay up to €5 for the book.

What price maximizes Readalot's profit? There are two natural prices to consider: €30 is the highest price Readalot can charge and still get the 100,000 diehard fans, and €5 is the highest price it can charge and still get the entire market of 500,000 potential readers. It is a matter of simple arithmetic to solve Readalot's problem. At a price of €30, Readalot sells 100,000 copies, has revenue of €3 million, and makes profit of €1 million. At a price of €5, it sells 500,000 copies, has revenue of €2.5 million, and makes a profit of €500,000. Thus, Readalot maximizes profit by charging €30 and forgoing the opportunity to sell to the 400,000 less enthusiastic readers.

Notice that Readalot's decision causes a deadweight loss. There are 400,000 readers willing to pay €5 for the book, and the marginal cost of providing it to them is zero. Thus, €2 million of total surplus is lost when Readalot charges the higher price. This deadweight loss is the usual inefficiency that arises whenever a monopolist charges a price above marginal cost.

Now suppose that Readalot's marketing department makes an important discovery: these two groups of readers are in separate markets. All the die-hard fans live in Switzerland and all the other readers live in

Turkey. Moreover, it is difficult for readers in one country to buy books in the other. How does this discovery affect Readalot's marketing strategy?

In this case, the company can make even more profit. To the 100,000 Swiss readers, it can charge €30 for the book. To the 400,000 Turkish readers, it can charge €5 for the book (or the Turkish lira equivalent). In this case, revenue is €3 million in Switzerland and €2 million in Turkey, for a total of €5 million. Profit is then €3 million, which is substantially greater than the €1 million the company could earn charging the same €30 price to all customers. Not surprisingly, Readalot chooses to follow this strategy of price discrimination.

Although the story of Readalot Publishing is hypothetical, it describes accurately the business practice of many publishing companies. Textbooks, for example, have been sold at different prices in Europe from that charged in the United States. Even more important is the price differential between hardcover books and paperbacks. When a publisher has a new novel, it initially releases an expensive hardcover edition and later releases a cheaper paperback edition. The difference in price between these two editions far exceeds the difference in printing costs. The publisher's goal is just as in our example. By selling the hardcover to die-hard fans (and libraries) who must have the book as soon as it is published, and the paperback to less enthusiastic readers who don't mind waiting, the publisher price discriminates and raises its profit.

The Moral of the Story

Like any parable, the story of Readalot Publishing is stylized. Yet, also like any parable, it teaches some important and general lessons. In this case, there are three lessons to be learned about price discrimination.

The first and most obvious lesson is that price discrimination is a rational strategy for a profit-maximizing monopolist. In other words, by charging different prices to different customers, a monopolist can increase its profit. In essence, a price-discriminating monopolist charges each customer a price closer to his or her willingness to pay than is possible with a single price.

The second lesson is that price discrimination requires the ability to separate customers according to their willingness to pay. In our example, customers were separated geographically. But sometimes monopolists choose other differences, such as age or income, to distinguish among customers. Energy companies are able to discriminate through setting different prices at different times of the day with off-peak usage priced lower than peak time. Similarly, rail companies charge different prices to passengers at certain times of the day with peak travel attracting a much higher price than off-peak travel. Where there is a difference in the price elasticity of demand the monopolist can exploit this and practice price discrimination. Between the hours of 6.00am and 9.30am on weekday mornings, for example, the price elasticity of demand for rail travel is relatively low whereas between 9.30am and 4.00pm it tends to be relatively high. A higher price can be charged at the peak time but during the off-peak period, the firm may benefit from charging a lower price and encouraging more passengers to travel; the cost of running the train is largely fixed and the marginal cost of carrying an additional passenger is almost zero. Lowering the price, therefore, is a way of utilizing the capacity on the train and adding to profit.

A corollary to this second lesson is that certain market forces can prevent firms from price discriminating. In particular, one such force is **arbitrage**, the process of buying a good in one market at a low price and selling it in another market at a higher price in order to profit from the price difference.

arbitrage the process of buying a good in one market at a low price and selling it in another market at a higher price in order to profit from the price difference

In our example, suppose that Swiss bookshops could buy the book in Turkey for €5 and resell it to Swiss readers at a price well below €30. This arbitrage would prevent Readalot from price discriminating because no Swiss resident would buy the book at the higher price. In fact, the increased use of the Internet for buying books and other goods through companies like Amazon is likely to affect the ability of companies to price discriminate internationally. Where firms can enforce the division of the market, as in the case of rail fares, it can practice price discrimination. A passenger buying a ticket at off-peak rates is not allowed to travel on a train running during peak periods, and hence arbitrage is circumvented.

The third lesson from our parable is perhaps the most surprising: price discrimination can raise economic welfare. Recall that a deadweight loss arises when Readalot charges a single €30 price, because the 400,000 less enthusiastic readers do not end up with the book, even though they value it at more than its marginal cost of production. By contrast, when Readalot price discriminates, all readers end up with the book, and the outcome is efficient. Thus, price discrimination can eliminate the inefficiency inherent in monopoly pricing.

Note that the increase in welfare from price discrimination shows up as higher producer surplus rather than higher consumer surplus. In our example, consumers are no better off for having bought the book: the price they pay exactly equals the value they place on the book, so they receive no consumer surplus. The entire increase in total surplus from price discrimination accrues to Readalot Publishing in the form of higher profit.

The Analytics of Price Discrimination

Let us consider a little more formally how price discrimination affects economic welfare. We begin by assuming that the monopolist can price discriminate perfectly. **Perfect price discrimination** describes a situation in which the monopolist knows exactly the willingness to pay of each customer and can charge each customer a different price. In this case, the monopolist charges each customer exactly his willingness to pay, and the monopolist gets the entire surplus in every transaction.

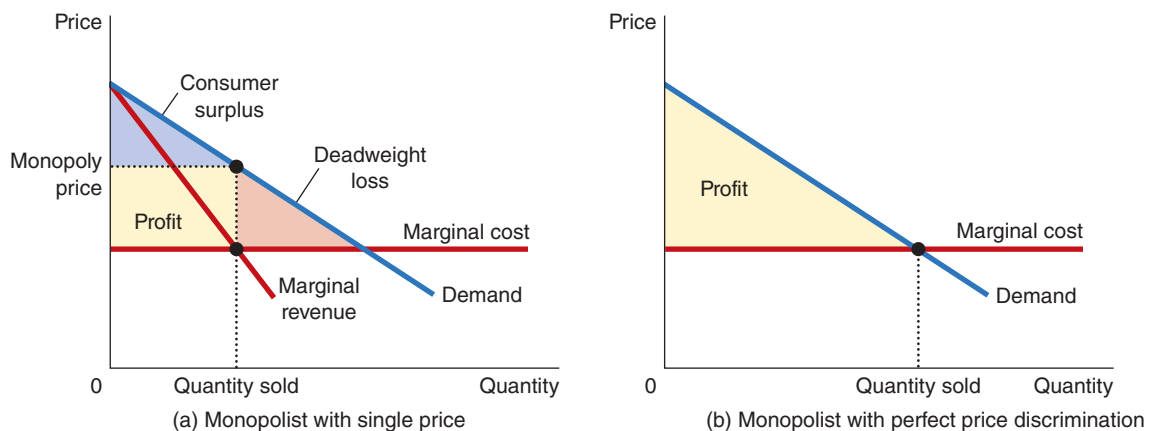
perfect price discrimination a situation in which the monopolist knows exactly the willingness to pay of each customer and can charge each customer a different price

Figure 14.9 shows producer and consumer surplus with and without price discrimination. Without price discrimination, the firm charges a single price above marginal cost, as shown in panel (a). Because some potential customers who value the good at more than marginal cost do not buy it at this high price, the monopoly causes a deadweight loss. Yet when a firm can perfectly price discriminate, as shown in panel (b), each customer who values the good at more than marginal cost buys the good and is charged his willingness to pay. All mutually beneficial trades take place, there is no deadweight loss, and the entire surplus derived from the market goes to the monopoly producer in the form of profit.

FIGURE 14.9

Welfare With and Without Price Discrimination

Panel (a) shows a monopolist that charges the same price to all customers. Total surplus in this market equals the sum of profit (producer surplus) and consumer surplus. Panel (b) shows a monopolist that can perfectly price discriminate. Because consumer surplus equals zero, total surplus now equals the firm's profit. Comparing these two panels, you can see that perfect price discrimination raises profit, raises total surplus and lowers consumer surplus.



In reality, of course, price discrimination is not perfect. Customers do not walk into shops with signs displaying their willingness to pay. Instead, firms price discriminate by dividing customers into groups: young versus old, weekday versus weekend shoppers, Germans versus British and so on. Unlike those in our parable of Readalot Publishing, customers within each group differ in their willingness to pay for the product, making perfect price discrimination impossible.

How does this imperfect price discrimination affect welfare? The analysis of these pricing schemes is quite complicated, and it turns out that there is no general answer to this question. Compared to the monopoly outcome with a single price, imperfect price discrimination can raise, lower or leave unchanged total surplus in a market. The only certain conclusion is that price discrimination raises the monopoly's profit – otherwise the firm would choose to charge all customers the same price.

Examples of Price Discrimination

Firms use various business strategies aimed at charging different prices to different customers. Now that we understand the economics of price discrimination, let's consider some examples.

Cinema Tickets Many cinemas charge a lower price for children and senior citizens than for other patrons. This fact is hard to explain in a competitive market. In a competitive market, price equals marginal cost, and the marginal cost of providing a seat for a child or senior citizen is the same as the marginal cost of providing a seat for anyone else. Yet this fact is easily explained if cinemas have some local monopoly power and if children and senior citizens have a lower willingness to pay for a ticket. In this case, cinemas raise their profit by price discriminating.

Airline Prices Seats on aeroplanes are sold at many different prices. Most airlines charge a lower price for a round-trip ticket between two cities if the traveller stays over a Saturday night. At first this seems odd. Why should it matter to the airline whether a passenger stays over a Saturday night? The reason is that this rule provides a way to separate business travellers and personal travellers. A passenger on a business trip has a high willingness to pay and, most likely, does not want to stay over a Saturday night. By contrast, a passenger travelling for personal reasons has a lower willingness to pay and is more likely to be willing to stay over a Saturday night. Thus, the airlines can successfully price discriminate by charging a lower price for passengers who stay over a Saturday night.

Discount Coupons Many companies offer discount coupons to the public in newspapers and magazines. A buyer simply has to cut out the coupon in order to get €0.50 off his next purchase. Why do companies offer these coupons? Why don't they just cut the price of the product by €0.50?

The answer is that coupons allow companies to price discriminate. Companies know that not all customers are willing to spend the time to cut out coupons. Moreover, the willingness to clip coupons is related to the customer's willingness to pay for the good. A rich and busy executive is unlikely to spend her time cutting discount coupons out of the newspaper, and she is probably willing to pay a higher price for many goods. A person who is unemployed is more likely to clip coupons and has a lower willingness to pay. Thus, by charging a lower price only to those customers who cut out coupons, firms can successfully price discriminate.

Quantity Discounts So far in our examples of price discrimination the monopolist charges different prices to different customers. Sometimes, however, monopolists price discriminate by charging different prices to the same customer for different units that the customer buys. Traditionally, English bakers would give you an extra cake for nothing if you bought 12. While the quaint custom of the 'baker's dozen' (i.e. 13 for the price of 12) is largely a thing of the past, many firms offer lower prices to customers who buy large quantities. This is a form of price discrimination because the customer effectively pays a higher price for the first unit bought than for the last. Quantity discounts are often a successful way of price discriminating because a customer's willingness to pay for an additional unit declines as the customer buys more units.

SELF TEST Give two examples of price discrimination. ● How does perfect price discrimination affect consumer surplus, producer surplus and total surplus?

PUBLIC POLICY TOWARDS MONOPOLIES

We have seen that monopolies, in contrast to competitive markets, fail to allocate resources efficiently. Monopolies produce less than the socially desirable quantity of output and, as a result, charge prices above marginal cost. Policymakers in the government can respond to the problem of monopoly in one of four ways, by:

- trying to make monopolized industries more competitive
- regulating the behaviour of the monopolies
- turning some private monopolies into public enterprises
- doing nothing at all.

All industrialized countries have some sort of process for legally prohibiting mergers or activity by firms with market power that are against the public interest. This is variously referred to as anti-trust law and anti-trust policy and also as competition law and competition policy depending on where in the world you are. These laws give governments various ways to promote competition.

In Europe, each country has a competition authority. In the UK it is the Competition Commission; in Germany it is the Federal Cartel Office (*Bundeskartellamt*); in 2009 the French Competition Authority began discharging its regulatory powers following reform of competition regulation; and in Italy the Anti-trust Authority (*Autorità garante della concorrenza e del mercato*) oversees competition issues. National competition authorities such as these cooperate with each other and with the European Union (EU) Competition Commission through the European Competition Network (ECN). The aim of the network is to coordinate activities and share information to help enforce EU competition law in member states where the opportunities for cross-border business have increased as the EU has developed and expanded.

Whilst each national country can enforce its own competition legislation, these laws have to be in line with overall EU competition legislation. In the UK, for example, the Competition Act 1998 and the Enterprise Act 2002 both deal with competition issues within the UK but cross-border competition cases would be dealt with under EU law. There are well-defined criteria for deciding whether a proposed merger of companies belonging to more than one EU country is subject to reference exclusively to the European Commission rather than to national authorities, such as the size of the worldwide or European turnover of the companies in question.

Competition legislation covers three main areas:

- Acting against cartels and cases where businesses engage in restrictive business practices which prevent free trade.
- Banning pricing strategies which are anti-competitive such as price fixing, predatory pricing, price gouging and so on, and through behaviour which might lead to a restriction in competition such as the sharing of information or carving up markets between different firms, rigging bids in tender processes or deliberately restricting production to reduce competition.
- Monitoring and supervising acquisitions and joint ventures.

The legislation allows competition authorities the right to fine firms who are found guilty of restricting competition, ordering firms to change behaviour and banning proposed acquisitions. The investigation will consider whether the acquisition, regardless of what size company it produces, is in the public interest. This is in recognition of the fact that companies sometimes merge not to reduce competition but to lower costs through more efficient joint production. These benefits from mergers are often called **synergies** where the perceived benefits of the combined operations are greater than those which would arise if the firms stayed separate.

synergies where the perceived benefits of the combined operations of a merged organization are greater than those which would arise if the firms stayed separate

Clearly, the government must be able to determine which mergers are desirable and which are not. That is, it must be able to measure and compare the social benefit from synergies to the social costs of reduced competition. In the UK, the Director-General of Fair Trading advises the Secretary of State for

Business Innovation and Skills (a government minister) on whether or not a particular merger should be referred for investigation by the Competition Commission – an independent body with members from private industry as well as some academic economists. The Competition Commission then reports its conclusions as to whether the proposed merger is in ‘the public interest’, so that the minister can rule on whether to allow or prohibit the merger (although the Secretary of State has the power to – and occasionally does – overrule the recommendations of the Competition Commission).

Regulation

Another way in which the government deals with the problem of monopoly is by regulating the behaviour of monopolists. This solution is common in the case of natural monopolies, such as utility companies like water, gas and electricity companies. These companies are not allowed to charge any price they want. Instead, government agencies regulate their prices.

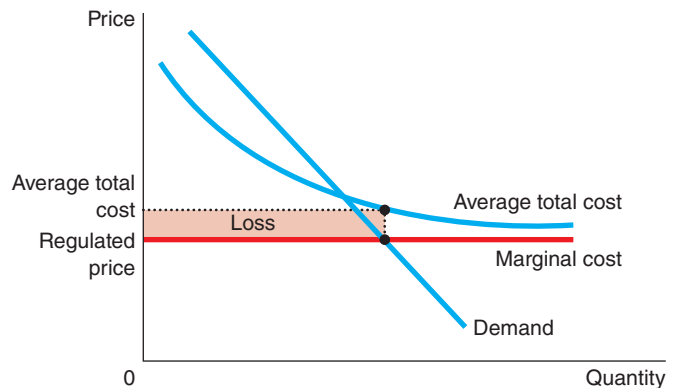
What price should the government set for a natural monopoly? This question is not as easy as it might at first appear. One might conclude that the price should equal the monopolist’s marginal cost. If price equals marginal cost, customers will buy the quantity of the monopolist’s output that maximizes total surplus, and the allocation of resources will be efficient.

There are, however, two practical problems with marginal-cost pricing as a regulatory system. The first is illustrated in Figure 14.10. Natural monopolies, by definition, have declining average total cost. When

FIGURE 14.10

Marginal Cost Pricing for a Natural Monopoly

Because a natural monopoly has declining average total cost, marginal cost is less than average total cost. Therefore, if regulators require a natural monopoly to charge a price equal to marginal cost, price will be below average total cost, and the monopoly will lose money.



average total cost is declining, marginal cost is less than average total cost. If regulators are to set price equal to marginal cost, that price will be less than the firm’s average total cost, and the firm will lose money. Instead of charging such a low price, the monopoly firm would just exit the industry.

Regulators can respond to this problem in various ways, none of which is perfect. One way is to subsidize the monopolist. In essence, the government picks up the losses inherent in marginal-cost pricing. Yet to pay for the subsidy, the government needs to raise money through taxation, which involves its own deadweight losses. Alternatively, the regulators can allow the monopolist to charge a price higher than marginal cost. If the regulated price equals average total cost, the monopolist earns exactly zero economic profit. Yet average-cost pricing leads to deadweight losses, because the monopolist’s price no longer reflects the marginal cost of producing the good. In essence, average-cost pricing is like a tax on the good the monopolist is selling.

The second problem with marginal-cost pricing as a regulatory system (and with average-cost pricing as well) is that it gives the monopolist no incentive to reduce costs. Each firm in a competitive market tries to reduce its costs because lower costs mean higher profits. But if a regulated monopolist knows that regulators will reduce prices whenever costs fall, the monopolist will not benefit from lower costs. In practice, regulators deal with this problem by allowing monopolists to keep some of the benefits from lower costs in the form of higher profit, a practice that requires some departure from marginal-cost pricing.

For example, in the UK, utility companies have often been subject to price caps whereby the regulator determines that the real price of the company's product – a kilowatt hour of electricity, for example – should fall by a given number of percentage points each year, reflecting productivity rises. Say, for example, this is 2 per cent. The company would then be allowed to raise its prices each year by the inflation rate *minus* 2 per cent. If the company increases its productivity by, say 4 per cent each year, however (in other words it can produce the same amount of output with 4 per cent less inputs), then in real terms its profits will go up each year. In this way, the system of price caps aims to give natural monopolies the motivation to improve efficiency and productivity that would be supplied by the invisible hand in a competitive market.

Public Ownership

The third policy used by the government to deal with monopoly is public ownership. That is, rather than regulating a natural monopoly that is run by a private firm, the government can run the monopoly itself. An industry owned by the government is called a nationalized industry. This solution is common in many European countries, where the government owns and operates utilities such as the telephone, water and electric companies.

Economists usually prefer private to public ownership of natural monopolies. The key issue is how the ownership of the firm affects the costs of production. Private owners have an incentive to minimize costs as long as they reap part of the benefit in the form of higher profit. If the firm's managers are doing a bad job of keeping costs down, the firm's owners will fire them. By contrast, if the government bureaucrats who run a monopoly do a bad job, the losers are the customers and taxpayers, whose only recourse is the political system. The bureaucrats may become a special interest group and attempt to block cost-reducing reforms. Put simply, as a way of ensuring that firms are well run, the voting booth is less reliable than the profit motive.

Doing Nothing

Each of the foregoing policies aimed at reducing the problem of monopoly has drawbacks. As a result, some economists argue that it is often best for the government not to try to remedy the inefficiencies of monopoly pricing. Here is the assessment of economist George Stigler, who won the Nobel Prize for his work in industrial organization, writing in the *Fortune Encyclopaedia of Economics*:

A famous theorem in economics states that a competitive enterprise economy will produce the largest possible income from a given stock of resources. No real economy meets the exact conditions of the theorem, and all real economies will fall short of the ideal economy – a difference called 'market failure'. In my view, however, the degree of 'market failure' for the American economy is much smaller than the 'political failure' arising from the imperfections of economic policies found in real political systems.

As this quotation makes clear, determining the proper role of the government in the economy requires judgements about politics as well as economics and the issues we discussed in the section on government failure in Chapter 11 have to be taken into account.

SELF TEST Describe the ways policymakers can respond to the inefficiencies caused by monopolies. List a potential problem with each of these policy responses.

CONCLUSION: THE PREVALENCE OF MONOPOLY

This chapter has discussed the behaviour of firms that have control over the prices they charge. We have seen that these firms behave very differently from the competitive firms studied in the previous chapter. Table 14.2 summarizes some of the key similarities and differences between competitive and monopoly markets.

From the standpoint of public policy, a crucial result is that monopolists produce less than the socially efficient quantity and charge prices above marginal cost. As a result, they cause deadweight losses. In

some cases, these inefficiencies can be mitigated through price discrimination by the monopolist, but at other times they call for policymakers to take an active role.

How prevalent are the problems of monopoly? There are two answers to this question.

In one sense, monopolies are common. Most firms have some control over the prices they charge. They are not forced to charge the market price for their goods, because their goods are not exactly the same as those offered by other firms. A Honda Accord is not the same as a Volkswagen Passat. Ben and Jerry's ice cream is not the same as Wall's. Each of these goods has a downwards sloping demand curve, which gives each producer some degree of monopoly power.

Yet firms with substantial monopoly power are quite rare. Few goods are truly unique. Most have substitutes that, even if not exactly the same are very similar. Ben and Jerry can raise the price of their ice cream a little without losing all their sales, but if they raise it very much sales will fall substantially.

In the end, monopoly power is a matter of degree. It is true that many firms have some monopoly power. It is also true that their monopoly power is usually limited. In these cases, we will not go far wrong assuming that firms operate in competitive markets, even if that is not precisely the case.

TABLE 14.2 Competition versus Monopoly: A Summary Comparison

| | Competition | Monopoly |
|--|------------------|------------------|
| Similarities | | |
| Goal of firms | Maximize profits | Maximize profits |
| Rule for maximizing | $MR = MC$ | $MR = MC$ |
| Can earn economic profits in the short run? | Yes | Yes |
| Differences | | |
| Number of firms | Many | One |
| Marginal revenue | $MR = P$ | $MR < P$ |
| Price | $P = MC$ | $P > MC$ |
| Produces welfare-maximizing level of output? | Yes | No |
| Entry in long run? | Yes | No |
| Can earn economic profits in long run? | No | Yes |
| Price discrimination possible? | No | Yes |

IN THE NEWS

Competition Law

Not every country in the world has strong competition laws to protect citizens from the potential for abuse of monopoly power. If strong laws do not exist will firms exploit their customers and even if they do will there always be an incentive to abuse monopoly power?

Competition Law in GCC Countries

As we have seen in this chapter, citizens in EU countries are protected from the potential for firms to abuse monopoly power by regulation and legislation. It is not the case everywhere in the world that such

protection exists. If there are no or limited government involvement in competition policy does this mean that firms with monopoly power will abuse that power?

Reports from one of the Gulf Cooperation Council (GCC) states suggests that there is a potential for this to happen. On a phone-in

to a radio show in Qatar one caller explained how he had bought a new Toyota but on reporting faults with the car to the dealer had not received satisfactory treatment and his complaint went unsolved. Other callers also complained of poor service from car dealers and others noted how they had felt they had been overcharged

by their telecom provider following a visit abroad but failed to get any satisfactory response.

One of the reasons for the complaints of exploitation by consumers is that the competition authorities in Qatar and the GCC do not have the sort of powers afforded their European counterparts. For example, electronics firms, including Panasonic, LG

Where competition law and regulation is weak, customers may find getting redress from firms for poor quality and/or service can be a frustrating experience.



and Philips, were fined €1.47 billion in 2012 for running cartels over a ten-year period to 2006 which fixed prices and divided up the markets for TV and computer monitors.

The GCC states (which consist of Saudi Arabia, Oman, United Arab Emirates, Kuwait and Bahrain) do have some competition laws; they are seen as being an important part of helping encourage foreign direct investment into the region. The laws that do exist, however, are seen as being piecemeal and not strong enough to prevent the sort of poor quality service experienced by callers to the radio programme in Qatar. As members of the World Trade Organization, however, there is an expectation that competition laws will be strengthened and a unified competition law for the GCC is in the process of being developed.

One aspect of strengthening the competition laws in the GCC is the development of associated legal services. If the concept of anti-trust is to be enshrined as part of the law in GCC states, then there has to be the legal expertise to not only help frame the law but also to interpret it. This may lead to greater cooperation between the GCC and other countries

especially given the complexities that arise when foreign companies are trading in GCC countries. This might explain why, for example, the complainant to the radio programme could not get help from Toyota's head office in Japan even though he took his argument to them.

Questions

- 1 What are the three main concerns that exist if a monopoly firm operates in a market without any regulation or legislation?**
- 2 Why do you think the firms in the article might have not given the sort of customer service the callers to the radio phone-in expected?**
- 3 Why do you think strong competition laws are important for attracting investment into countries such as the members of the GCC?**
- 4 To what extent are competition laws dependent on the legal system?**
- 5 Does the fact that the EU Competition Commission levied a record fine against electronics companies in December 2012 suggest that even if competition laws exist firms are likely to abuse monopoly power?**

SUMMARY

- A monopoly is a firm that is the sole seller in its market. A monopoly arises when a single firm owns a key resource, when the government gives a firm the exclusive right to produce a good, or when a single firm can supply the entire market at a smaller cost than many firms could.
- Because a monopoly is the sole producer in its market, it faces a downwards sloping demand curve for its product. When a monopoly increases production by 1 unit, it causes the price of its good to fall, which reduces the amount of revenue earned on all units produced. As a result, a monopoly's marginal revenue is always below the price of its good.
- Like a competitive firm, a monopoly firm maximizes profit by producing the quantity at which marginal revenue equals marginal cost. The monopoly then chooses the price at which that quantity is demanded. Unlike a competitive firm, a monopoly firm's price exceeds its marginal revenue, so its price exceeds marginal cost.
- A monopolist's profit-maximizing level of output is below the level that maximizes the sum of consumer and producer surplus. That is, when the monopoly charges a price above marginal cost, some consumers who value the good more than its cost of production do not buy it. As a result, monopoly causes deadweight losses similar to the deadweight losses caused by taxes.

- Policymakers can respond to the inefficiency of monopoly behaviour in four ways. They can use competition law to try to make the industry more competitive. They can regulate the prices that the monopoly charges. They can turn the monopolist into a government-run enterprise. Or, if the market failure is deemed small compared to the inevitable imperfections of policies, they can do nothing at all.
- Monopolists often can raise their profits by charging different prices for the same good based on a buyer's

willingness to pay. This practice of price discrimination can raise economic welfare by getting the good to some consumers who otherwise would not buy it. In the extreme case of perfect price discrimination, the deadweight losses of monopoly are completely eliminated. More generally, when price discrimination is imperfect, it can either raise or lower welfare compared to the outcome with a single monopoly price.

QUESTIONS FOR REVIEW

- 1 What are barriers to entry? What are the main barriers to entry?
- 2 What are the main sources of monopoly power?
- 3 Give an example of a government-created monopoly. Is creating this monopoly necessarily bad public policy? Explain.
- 4 Define natural monopoly. What does the size of a market have to do with whether an industry is a natural monopoly?
- 5 Why is a monopolist's marginal revenue less than the price of its good? Can marginal revenue ever be negative? Explain.
- 6 Draw the demand, marginal revenue and marginal cost curves for a monopolist. Show the profit-maximizing level of output. Show the profit-maximizing price.
- 7 In your diagram from the previous question, show the level of output that maximizes total surplus. Show the deadweight loss from the monopoly. Explain your answer.
- 8 What gives the government the power to regulate mergers between firms? From the standpoint of the welfare of society, give a good reason and a bad reason that two firms might want to merge.
- 9 Describe the two problems that arise when regulators tell a natural monopoly that it must set a price equal to marginal cost.
- 10 Give two examples of price discrimination. In each case, explain why the monopolist chooses to follow this business strategy.

PROBLEMS AND APPLICATIONS

- 1 A publisher faces the following demand schedule for the next novel of one of its popular authors:

| Price (€) | Quantity demanded |
|-----------|-------------------|
| 100 | 0 |
| 90 | 100,000 |
| 80 | 200,000 |
| 70 | 300,000 |
| 60 | 400,000 |
| 50 | 500,000 |
| 40 | 600,000 |
| 30 | 700,000 |
| 20 | 800,000 |
| 10 | 900,000 |
| 0 | 1,000,000 |

The author is paid €2 million to write the book, and the marginal cost of publishing the book is a constant €10 per book.

- a. Compute total revenue, total cost and profit at each quantity. What quantity would a profit-maximizing publisher choose? What price would it charge?

- b. Compute marginal revenue. (Recall that $MR = \Delta TR / \Delta Q$.) How does marginal revenue compare to the price? Explain.
- c. Graph the marginal revenue, marginal cost and demand curves. At what quantity do the marginal revenue and marginal cost curves cross? What does this signify?
- d. In your graph, shade in the deadweight loss. Explain in words what this means.
- e. If the author were paid €3 million instead of €2 million to write the book, how would this affect the publisher's decision regarding the price to charge? Explain.
- f. Suppose the publisher was not profit-maximizing but was concerned with maximizing economic efficiency. What price would it charge for the book? How much profit would it make at this price?

- 2 Consider the delivery of mail. In general, what is the shape of the average total cost curve? How might the shape differ between isolated rural areas and densely populated urban areas? How might the shape have changed over time? Explain.

- 3 Suppose the Eau de Jeunesse Water Company has a monopoly on bottled water sales in France. If the price of tap water increases, what is the change in Eau de Jeunesse's profit-maximizing levels of output, price and profit? Explain in words and with a graph.
- 4 A small town is served by many competing supermarkets, which have constant marginal cost.
- Using a diagram of the market for groceries, show the consumer surplus, producer surplus and total surplus.
 - Now suppose that the independent supermarkets combine into one chain. Using a new diagram, show the new consumer surplus, producer surplus and total surplus. Relative to the competitive market, what is the transfer from consumers to producers? What is the deadweight loss?
- 5 A company is considering building a bridge across a river. The bridge would cost €2 million to build and nothing to maintain. The following table shows the company's anticipated demand over the lifetime of the bridge:

| Price per crossing (€) | Number of crossings (in thousands) |
|------------------------|------------------------------------|
| 8 | 0 |
| 7 | 100 |
| 6 | 200 |
| 5 | 300 |
| 4 | 400 |
| 3 | 500 |
| 2 | 600 |
| 1 | 700 |
| 0 | 800 |

- If the company was to build the bridge, what would be its profit-maximizing price? Would that be the efficient level of output? Why or why not?
 - If the company is interested in maximizing profit, should it build the bridge? What would be its profit or loss?
 - If the government were to build the bridge, what price should it charge for passengers and vehicles to use the bridge? Explain your answer.
 - Should the government build the bridge? Explain.
- 6 The Placebo Drug Company holds a patent on one of its discoveries.
- Assuming that the production of the drug involves rising marginal cost, draw a diagram to illustrate Placebo's profit-maximizing price and quantity. Also show Placebo's profits.
 - Now suppose that the government imposes a tax on each bottle of the drug produced. On a new diagram, illustrate Placebo's new price and quantity. How does each compare to your answer in part (a)?
 - Although it is not easy to see in your diagrams, the tax reduces Placebo's profit. Explain why this must be true.
- Instead of the tax per bottle, suppose that the government imposes a tax on Placebo of €110,000 regardless of how many bottles are produced. How does this tax affect Placebo's price, quantity and profits? Explain.
- 7 Pablo, Dirk and Franz run the only saloon in town. Pablo wants to sell as many drinks as possible without losing money. Dirk wants the saloon to bring in as much revenue as possible. Franz wants to make the largest possible profits. Using a single diagram of the saloon's demand curve and its cost curves, show the price and quantity combinations favoured by each of the three partners. Explain.
- 8 The Best Computer Company just developed a new computer chip, on which it immediately acquires a patent.
- Draw a diagram that shows the consumer surplus, producer surplus and total surplus in the market for this new chip.
 - What happens to these three measures of surplus if the firm can perfectly price discriminate? What is the change in deadweight loss? What transfers occur?
- 9 Explain why a monopolist will always produce a quantity at which the demand curve is elastic. (Hint: if demand is inelastic and the firm raises its price, what happens to total revenue and total costs?)
- 10 Many schemes for price discriminating involve some cost. For example, discount coupons take up time and resources from both the buyer and the seller. This question considers the implications of costly price discrimination. To keep things simple, let's assume that our monopolist's production costs are simply proportional to output, so that average total cost and marginal cost are constant and equal to each other.
- Draw the cost, demand and marginal revenue curves for the monopolist. Show the price the monopolist would charge without price discrimination.
 - In your diagram, mark the area equal to the monopolist's profit and call it X. Mark the area equal to consumer surplus and call it Y. Mark the area equal to the deadweight loss and call it Z.
 - Now suppose that the monopolist can perfectly price discriminate. What is the monopolist's profit? (Give your answer in terms of X, Y and Z.)
 - What is the change in the monopolist's profit from price discrimination? What is the change in total surplus from price discrimination? Which change is larger? Explain. (Give your answer in terms of X, Y and Z.)
 - Now suppose that there is some cost of price discrimination. To model this cost, let's assume that the monopolist has to pay a fixed cost C in order to price discriminate. How would a monopolist make the decision whether to pay this fixed cost? (Give your answer in terms of X, Y, Z and C .)

15

MARKET STRUCTURES II: MONOPOLISTIC COMPETITION

You arrive at university and set off into the town for a taste of the night life. The town has eight night clubs all in walking distance of the main transport hub. Each one has music, bars, light snacks, places to talk, different prices for food and drink, different entry prices, rules and closing times. When you are choosing between these different night clubs what kind of market are you participating in?

On the one hand, the market for night clubs seems competitive. In most towns and cities there are plenty of clubs vying for your attention. A buyer in this overall market has thousands of competing products from which to choose. And because entrepreneurs are willing to put the time and effort into setting up and running night clubs, there are a lot of people wanting to enter the market.

On the other hand, the market for night clubs seems monopolistic. Because each club is able to present itself in a unique way, night club owners have some latitude in choosing what price to charge. The sellers in this market are price makers rather than price takers. The price of entry into a night club greatly exceeds the marginal cost of one extra person entering.

In this chapter we examine markets that have some features of competition and some features of monopoly. This market structure is called **monopolistic competition**. Monopolistic competition describes a market with the following attributes:

monopolistic competition a market structure in which many firms sell products that are similar but not identical

- *Many sellers.* There are many firms competing for the same group of customers.
- *Product differentiation.* Each firm produces a product that is at least slightly different from those of other firms, whether physically different or whether perceived as being different by consumers. The firm is able to have some control over the extent to which it can differentiate its product from its rivals thus reducing the degree of substitutability and garnering an element of customer or brand loyalty. Therefore, rather than being a price taker, each firm faces a downwards sloping demand curve.
- *Free entry.* Firms can enter (or exit) the market without restriction. Thus, the number of firms in the market adjusts until economic profits are driven to zero.

Table 15.1 lists some other examples of the types of market with these attributes, in addition to the example of night clubs we have opened this chapter with.

Monopolistic competition is a market structure that lies between the extreme cases of competition and monopoly. Under monopolistic competition, there are many sellers, each of which is small compared to the market. A monopolistically competitive market departs from the perfectly competitive ideal because each of the sellers is able to offer a different product by differentiating it in some way from other suppliers. As a result, it is another example of imperfect competition.

TABLE 15.1**Examples of Markets which have Characteristics of Monopolistic Competition**

| | |
|---------------------------|--------------------------|
| Computer games | Vets |
| Restaurants | Hotel accommodation |
| Conference organizers | Air conditioning systems |
| Wedding planners | Pest control |
| Plumbing | Removal services |
| Coach hire | Beauty consultants |
| Funeral directors | Shop fitters |
| Fabric manufacturers | Waste disposal |
| Tailors | Dentists |
| Music teachers | Children's entertainers |
| Books | Gas engineers |
| CDs/DVDs | Steel fabricators |
| Landscape architects | Driving schools |
| Environmental consultants | Opticians |
| Furniture manufacturers | Chimney sweeps |

COMPETITION WITH DIFFERENTIATED PRODUCTS

To understand monopolistically competitive markets, we first consider the decisions facing an individual firm. We then examine what happens in the long run as firms enter and exit the industry. Next, we compare the equilibrium under monopolistic competition to the equilibrium under perfect competition. Finally, we consider whether the outcome in a monopolistically competitive market is desirable from the standpoint of society as a whole.

The Monopolistically Competitive Firm in the Short Run

Each firm in a monopolistically competitive market is, in many ways, like a monopoly. Because its product is different from those offered by other firms, it faces a downwards sloping demand curve. (By contrast, remember, a perfectly competitive firm faces a horizontal demand curve at the market price.) Thus, the monopolistically competitive firm follows a monopolist's rule for profit maximization: it chooses the quantity at which marginal revenue equals marginal cost and then uses its demand curve to find the price consistent with that quantity.

Figure 15.1 shows the cost, demand and marginal revenue curves for two typical firms, each in a different monopolistically competitive industry. In both panels of this figure, the profit-maximizing quantity is found at the intersection of the marginal revenue and marginal cost curves. The two panels in this figure show different outcomes for the firm's profit. In panel (a), price exceeds average total cost, so the firm makes a profit. In panel (b), price is below average total cost. In this case, the firm is unable to make a positive profit, so the best the firm can do is to minimize its losses.

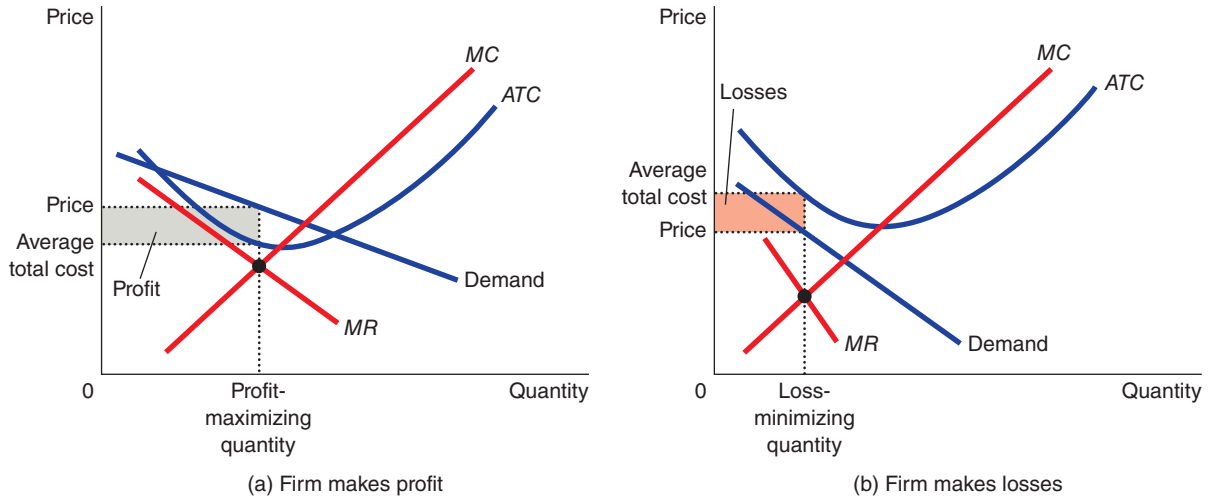
All this should seem familiar. A monopolistically competitive firm chooses its quantity and price just as a monopoly does. In the short run, these two types of market structure are similar.

The Long-Run Equilibrium

The situations depicted in Figure 15.1 do not last long. When demand seems to be strong and firms are making profits, as in panel (a), new firms have an incentive to enter the market (remember that there is free entry and exit into the market). This entry means that more firms are now offering products for sale in the industry. For example, the popularity of night clubs has led to more people wanting to set up new night clubs to take advantage of the demand and the profits that can be made. The increase in supply means customers have more choice and causes the price received by all firms in the industry to fall. If an existing firm wishes to sell more then it must reduce its price. There are now more substitutes available in the market and so the effect for individual firms is that the demand curve for its product shifts to the left as the number of products from which customers can now choose increases and, therefore, reduces the demand faced

FIGURE 15.1**Monopolistic Competitors in the Short Run**

Monopolistic competitors, like monopolists, maximize profit by producing the quantity at which marginal revenue equals marginal cost. The firm in panel (a) makes a profit because, at this quantity, price is above average total cost. The firm in panel (b) makes losses because, at this quantity, price is less than average total cost.



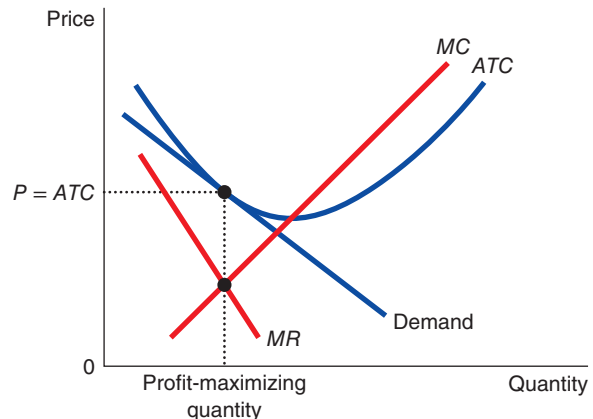
by each firm already in the market. Profit encourages entry, and entry shifts the demand curves faced by firms to the left. As the demand for firms' products falls, these firms experience declining profit.

Some firms in the industry will have just been surviving, but as new firms enter and their demand curve shifts to the left they may find themselves making sub-normal profits and as a result might decide to leave the industry. When firms are making losses, as in panel (b), firms in the market have an incentive to exit. As firms exit, the supply will fall and price will rise. There are now fewer substitutes and so customers have fewer products from which to choose. This decrease in the number of firms effectively expands the demand faced by those firms that stay in the market. In other words, losses encourage exit, and exit has the effect of shifting the demand curves of the remaining firms to the right. As the demand for the remaining firms' products rises, these firms experience rising profit (that is, declining losses).

This process of entry and exit continues until the firms in the market are making exactly zero economic profit (normal profit). Figure 15.2 depicts the long-run equilibrium. Once the market reaches this equilibrium, new firms have no incentive to enter, and existing firms have no incentive to exit.

FIGURE 15.2**A Monopolistic Competitor in the Long Run**

In a monopolistically competitive market, if firms are making profit, new firms enter and the demand curves for the incumbent firms shift to the left. Similarly, if firms are making losses, old firms exit and the demand curves of the remaining firms shift to the right. Because of these shifts in demand, a monopolistically competitive firm eventually finds itself in the long-run equilibrium shown here. In this long-run equilibrium, price equals average total cost, and the firm earns zero profit.



Notice that the demand curve in this figure just barely touches the average total cost curve. Mathematically, we say the two curves are *tangential* to each other. These two curves must be tangential once entry and exit have driven profit to zero. Because profit per unit sold is the difference between price (found on the demand curve) and average total cost, the maximum profit is zero only if these two curves touch each other without crossing.

To sum up, two characteristics describe the long-run equilibrium in a monopolistically competitive market:

- As in a monopoly market, price exceeds marginal cost. This conclusion arises because profit maximization requires marginal revenue to equal marginal cost and because the downwards sloping demand curve makes marginal revenue less than the price.
- As in a competitive market, price equals average total cost. This conclusion arises because free entry and exit drive economic profit to zero.

The second characteristic shows how monopolistic competition differs from monopoly. Because a monopoly is the sole seller of a product without close substitutes, it can earn positive economic profit, even in the long run. By contrast, because there is free entry into a monopolistically competitive market, the economic profit of a firm in this type of market is driven to zero.

Monopolistic versus Perfect Competition

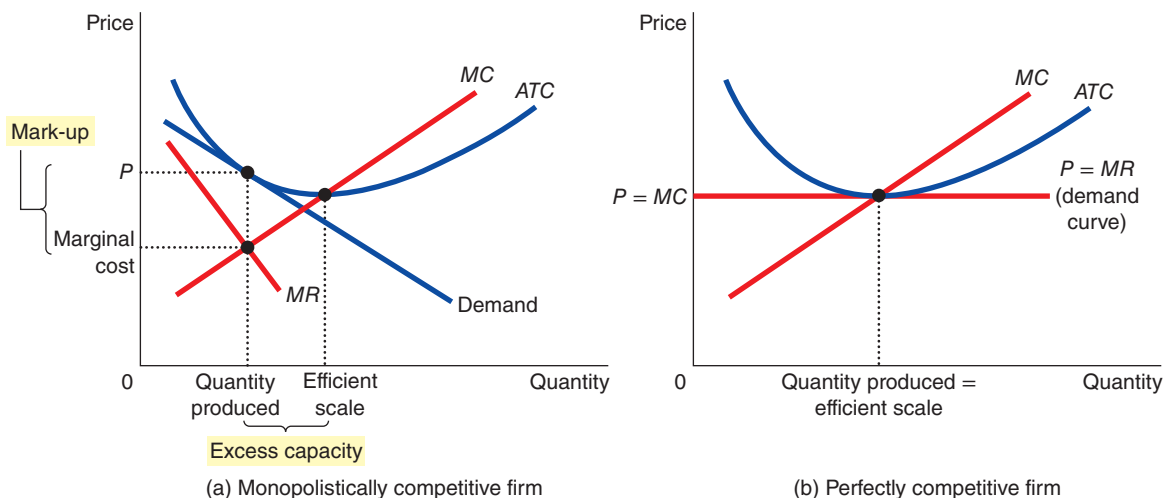
Figure 15.3 compares the long-run equilibrium under monopolistic competition to the long-run equilibrium under perfect competition. There are two noteworthy differences between monopolistic and perfect competition – excess capacity and the mark-up.

Excess Capacity As we have just seen, entry and exit drive each firm in a monopolistically competitive market to a point of tangency between its demand and average total cost curves. Panel (a) of Figure 15.3 shows that the quantity of output at this point is smaller than the quantity that minimizes average total

FIGURE 15.3

Monopolistic versus Perfect Competition

Panel (a) shows the long-run equilibrium in a monopolistically competitive market, and panel (b) shows the long-run equilibrium in a perfectly competitive market. Two differences are notable. (1) The perfectly competitive firm produces at the efficient scale, where average total cost is minimized. By contrast, the monopolistically competitive firm produces at less than the efficient scale. (2) Price equals marginal cost under perfect competition, but price is above marginal cost under monopolistic competition.



cost. Thus, under monopolistic competition, firms produce on the downwards sloping portion of their average total cost curves. In this way, monopolistic competition contrasts starkly with perfect competition. As panel (b) of Figure 15.3 shows, free entry in competitive markets drives firms to produce at the minimum of average total cost.

The quantity that minimizes average total cost is called the *efficient scale* of the firm. In the long run, perfectly competitive firms produce at the efficient scale, whereas monopolistically competitive firms produce below this level. Firms are said to have *excess capacity* under monopolistic competition. In other words, a monopolistically competitive firm, unlike a perfectly competitive firm, could increase the quantity it produces and lower the average total cost of production.

Mark-Up over Marginal Cost A second difference between perfect competition and monopolistic competition is the relationship between price and marginal cost. For a competitive firm, such as that shown in panel (b) of Figure 15.3, price equals marginal cost. For a monopolistically competitive firm, such as that shown in panel (a), price exceeds marginal cost, because the firm always has some market power.

How is this mark-up over marginal cost consistent with free entry and zero profit? The zero-profit condition ensures only that price equals average total cost. It does *not* ensure that price equals marginal cost. Indeed, in the long-run equilibrium, monopolistically competitive firms operate on the declining portion of their average total cost curves, so marginal cost is below average total cost. Thus, for price to equal average total cost, price must be above marginal cost.

In this relationship between price and marginal cost, we see a key behavioural difference between perfect competitors and monopolistic competitors. Imagine that you were to ask a firm the following question: 'Would you like to see another customer come through your door ready to buy from you at your current price?' A perfectly competitive firm would answer that it didn't care. Because price exactly equals marginal cost, the profit from an extra unit sold is zero. By contrast, a monopolistically competitive firm is always eager to get another customer. Because its price exceeds marginal cost, an extra unit sold at the posted price means more profit. According to an old quip, monopolistically competitive markets are those in which sellers send greetings cards to the buyers.

Monopolistic Competition and the Welfare of Society

Is the outcome in a monopolistically competitive market desirable from the standpoint of society as a whole? Can policymakers improve on the market outcome? There are no simple answers to these questions.

One source of inefficiency is the mark-up of price over marginal cost. Because of the mark-up, some consumers who value the good at more than the marginal cost of production (but less than the price) will be deterred from buying it. Thus, a monopolistically competitive market has the normal deadweight loss of monopoly pricing. We first saw this type of inefficiency when we discussed monopoly in Chapter 14.

Although this outcome is clearly undesirable compared to the first-best outcome of price equal to marginal cost, there is no easy way for policymakers to fix the problem. To enforce marginal-cost pricing, policymakers would need to regulate all firms that produce differentiated products. Because such products are so common in the economy, the administrative burden of such regulation would be overwhelming.

Moreover, regulating monopolistic competitors would entail all the problems of regulating natural monopolies. In particular, because monopolistic competitors are making zero profits already, requiring them to lower their prices to equal marginal cost would cause them to make losses. To keep these firms in business, the government would need to help them cover these losses. Rather than raising taxes to pay for these subsidies, policymakers may decide it is better to live with the inefficiency of monopolistic pricing.

Another way in which monopolistic competition may be socially inefficient is that the number of firms in the market may not be the 'ideal' one. That is, there may be too much or too little entry. One way to think about this problem is in terms of the externalities associated with entry. Whenever a new firm considers

entering the market with a new product, it considers only the profit it would make. Yet its entry would also have two external effects:

- *The product-variety externality.* Because consumers get some consumer surplus from the introduction of a new product, entry of a new firm conveys a positive externality on consumers.
- *The business-stealing externality.* Because other firms lose customers and profits from the entry of a new competitor, entry of a new firm imposes a negative externality on existing firms.

Thus, in a monopolistically competitive market, there are both positive and negative externalities associated with the entry of new firms. Depending on which externality is larger, a monopolistically competitive market could have either too few or too many products.

Both of these externalities are closely related to the conditions for monopolistic competition. The product-variety externality arises because a new firm would offer a product different from those of the existing firms. The business-stealing externality arises because firms post a price above marginal cost and, therefore, are always eager to sell additional units. Conversely, because perfectly competitive firms produce identical goods and charge a price equal to marginal cost, neither of these externalities exists under perfect competition.

In the end, we can conclude only that monopolistically competitive markets do not have all the desirable welfare properties of perfectly competitive markets. That is, the invisible hand does not ensure that total surplus is maximized under monopolistic competition. Yet because the inefficiencies are subtle, hard to measure and hard to fix, there is no easy way for public policy to improve the market outcome.

SELF TEST List the three key attributes of monopolistic competition. ● Draw and explain a diagram to show the long-run equilibrium in a monopolistically competitive market. How does this equilibrium differ from that in a perfectly competitive market?

ADVERTISING AND BRANDING

It is nearly impossible to go through a typical day in a modern economy without being bombarded with advertising or to ignore the existence of brand names. Whether you are reading a newspaper, watching television or driving your car, some firm will try to convince you to buy its product or put its brand in front of you. Such behaviour is a natural feature of monopolistic competition. When firms sell differentiated products and charge prices above marginal cost, each firm has an incentive to advertise or develop brands in order to attract more buyers to its particular product or develop loyalty.

The amount of advertising varies substantially across products. Firms that sell highly differentiated consumer goods, such as over-the-counter drugs, perfumes, soft drinks, razor blades, breakfast cereals and dog food, typically spend between 10 and 20 per cent of revenue for advertising. Firms that sell industrial products, such as drill presses and communications satellites, typically spend very little on advertising. And firms that sell homogeneous products, such as wheat, peanuts or crude oil, spend nothing at all. Estimates from Nielsen suggest that firms such as Procter and Gamble and British Sky Broadcasting (BskyB) spent over £10 million (€12.3 million) on advertising in the UK alone and eMarketer reported that total media advertising spend in 2012 was around €397 billion. Advertising takes many forms: newspapers, magazines, TV, radio and via the Internet, a medium which is becoming increasingly important.

The Debate over Advertising

Is society wasting the resources it devotes to advertising? Or does advertising serve a valuable purpose? Assessing the social value of advertising is difficult and often generates heated argument among economists. Let's consider both sides of the debate.

The Critique of Advertising Critics of advertising argue that firms advertise in order to manipulate people's tastes. Much advertising is psychological rather than informational. Consider, for example, the

CASE STUDY Advertising – What Does it Really Do?

Ask many people the question ‘why do firms advertise?’ and they are likely to tell you that it is an attempt by firms to try and increase demand for their products or services. If you consider this view intuitively it might make sense but then ask yourself the question, ‘how many times have you seen an advert on the TV and then rushed out to buy the product advertised?’ The chances are that this has not (consciously) happened very often at all. So if adverts do not make us rush out to buy products what do they do?

Sutherland and Sylvester (2000) argue that it is largely a myth that adverts are designed to persuade us to buy products or services. They point out the following:

Advertising influences the order in which we evoke or notice the alternatives we consider. This does not feel like persuasion and it is not. It is nevertheless effective. Instead of persuasion and other major effects we should look for ‘feathers’, or minor effects. These can tip the balance when alternative brands are otherwise equal and, through repetition, can grow imperceptibly by small increments over time.

They liken the effect of advertising to that of watching someone grow up. You know that they are growing but the day-to-day changes in the individual are imperceptible. If you have not seen someone for some time, however, you do tend to notice the difference in their height, shape, features and so on. So it is with many advertising campaigns. The primary aim, they argue, is to generate a series of small effects, which ultimately influence our behaviour and may cause us to view differently the products or the brands that we choose, especially in a crowded marketplace with a large amount of competition.

Exactly how adverts work, therefore, is not easy to quantify. Sutherland and Sylvester suggest that many involved in the advertising industry do not really understand why some adverts seem to work and others don’t work anything like as well. It has long been recognized that psychology has a lot to do with advertising. Our understanding of the way the brain works has been revolutionized by the developments afforded by magnetic resonance imaging (MRI) scans. The advertising industry has not been slow in looking at this technology and its potential for improving the focus and efficiency of advertising.

In essence, this technique looks at the response of the brain to different images and messages. Using MRI techniques, the areas of the brain that respond to different stimuli can be identified. The field developed as a result of work carried out by a neuroscientist called Read Montague. Montague is Professor at the VTC Research Institute at Virginia Tech in the United States. Whilst in a previous position at Baylor College of Medicine in Texas, he presided over a challenge to see if people preferred different brands of cola. His initial studies suggested that people preferred Pepsi but their buying behaviour tended to favour Coke. In a repeat of the experiment in 2004 using MRI technology, Read and his colleagues gave a group of individuals two colas, Pepsi and Coke, to taste and asked them to state which they preferred. The respondents did not know that the two colas were in fact Pepsi and Coke. The subjects stated preferences were almost 50:50 although activity in a part of the brain associated with processing feelings of reward showed a higher response for Pepsi.

However, when the experiment was repeated and the respondents were told what they were drinking, around 75 per cent stated that they preferred Coke. Montague found that brain activity in the lateral pre-frontal cortex showed signs of enhanced activity during the exercise. This area of the brain is associated with higher level thinking. Montague posited that the brain was now using memories and making an association with the images and messages associated with commercials for Coke that respondents had witnessed over the years. He also suggested that such activity might lead to consumers preferring one product to another, even if there was other evidence to suggest that under normal circumstances, they would not have chosen that product.

Could you tell the difference between Pepsi and Coca-Cola? A lot of people think they can but maybe we are influenced by more than just taste.



References: Sutherland, M. and Sylvester, A.K. (2000) *Advertising and the Mind of the Consumer: What Works, What Doesn't, and Why*. St Leonards, New South Wales, Allen and Unwin.

typical television advert for some brand of soft drink. The advert most likely does not tell the viewer about the product's price or quality. Instead, it might show a group of happy people at a party on a beach on a beautiful sunny day. In their hands are cans of the soft drink. The goal of the advert is to convey a subconscious (if not subtle) message: 'You too can have many friends and be happy and beautiful, if only you drink our product.' Critics of advertising argue that such an advert creates a desire that otherwise might not exist.

Critics also argue that advertising impedes competition. Advertising often tries to convince consumers that products are more different than they truly are. By increasing the perception of product differentiation and fostering brand loyalty, advertising makes buyers less concerned with price differences among similar goods. With a less elastic demand curve, each firm charges a larger mark-up over marginal cost.

The Defence of Advertising Defenders of advertising argue that firms use advertising to provide information to customers. Advertising conveys the prices of the goods being offered for sale, the existence of new products and the locations of retail outlets. This information allows customers to make better choices about what to buy and, thus, enhances the ability of markets to allocate resources efficiently.

Defenders also argue that advertising fosters competition. Because advertising allows customers to be more fully informed about all the firms in the market, customers can more easily take advantage of price differences. Thus, each firm has less market power. In addition, advertising allows new firms to enter more easily, because it gives entrants a means to attract customers from existing firms.

Advertising as a Signal of Quality

Many types of advertising contain little apparent information about the product being advertised. Consider a firm introducing a new breakfast cereal. A typical advertisement might have some highly paid actor eating the cereal and exclaiming how wonderful it tastes. How much information does the advertisement really provide?

The answer is: more than you might think. Defenders of advertising argue that even advertising that appears to contain little hard information may in fact tell consumers something about product quality. The willingness of the firm to spend a large amount of money on advertising can itself be a *signal* to consumers about the quality of the product being offered.

Consider a hypothetical problem facing Nestlé and Kellogg. Each company has just come up with a recipe for a new breakfast cereal, which it would sell for €3 a box. To keep things simple, let's assume that the marginal cost of making cereal is zero, so the €3 is all profit. Each company knows that if it spends €10 million on advertising, it will get 1 million consumers to try its new cereal. And each company knows that if consumers like the cereal, they will buy it not once but many times.

First, consider Kellogg's decision. Based on market research, Kellogg knows that its cereal is only mediocre. Although advertising would sell one box to each of 1 million consumers, the consumers would quickly learn that the cereal is not very good and stop buying it. Kellogg decides it is not worth paying €10 million in advertising to get only €3 million in sales. So it does not bother to advertise. It sends its cooks back to the drawing board to find another recipe.

Nestlé, on the other hand, knows that its cereal is great. Each person who tries it will buy a box a month for the next year. Thus, the €10 million in advertising will bring in €36 million in sales. Advertising is profitable here because Nestlé has a good product that consumers will buy repeatedly. Thus, Nestlé chooses to advertise.

Now that we have considered the behaviour of the two firms, let's consider the behaviour of consumers. We began by asserting that consumers are inclined to try a new cereal that they see advertised. But is this behaviour rational? Should a consumer try a new cereal just because the seller has chosen to advertise it?

In fact, it may be completely rational for consumers to try new products that they see advertised. In our story, consumers decide to try Nestlé's new cereal because Nestlé advertises. Nestlé chooses to advertise because it knows that its cereal is quite good, while Kellogg chooses not to advertise because it knows that its cereal is only mediocre. By its willingness to spend money on advertising, Nestlé signals to

consumers the quality of its cereal. Each consumer thinks, quite sensibly, 'If the Nestlé Company is willing to spend so much money advertising this new cereal, it must be really good.'

What is most surprising about this theory of advertising is that the content of the advertisement is irrelevant. Nestlé signals the quality of its product by its willingness to spend money on advertising. (This example is used for illustrative purposes only and is not meant to infer that Kellogg deliberately produces inferior products!)

What the advertisements say is not as important as the fact that consumers know ads are expensive. By contrast, cheap advertising cannot be effective at signalling quality to consumers. In our example, if an advertising campaign cost less than €3 million, both Nestlé and Kellogg would use it to market their new cereals. Because both good and mediocre cereals would be advertised, consumers could not infer the quality of a new cereal from the fact that it is advertised. Over time, consumers would learn to ignore such cheap advertising.

This theory can explain why firms pay celebrities large amounts of money to make advertisements that, on the surface, appear to convey no information at all. The information is not in the advertisement's content, but simply in its existence and expense.

Branding and Brand Names

Advertising is closely related to the existence of **branding**. In many markets, there are two types of firms. Some firms sell products with widely recognized brand names, while other firms sell generic substitutes. For example, in a typical supermarket, you can find Pepsi next to less familiar colas, or Kellogg's cornflakes next to the supermarket's own brand of cornflakes, made for it by an unknown firm. Most often, the firm with the famous brand name spends more on advertising and charges a higher price for its product.

branding the means by which a business creates an identity for itself and highlights the way in which it differs from its rivals

Just as there is disagreement about the economics of advertising, there is disagreement about the economics of brand names and branding. Let's consider both sides of the debate.

Critics of brand names argue that branding causes consumers to perceive differences that do not really exist. In many cases, the generic good is almost indistinguishable from the brand-name good. Consumers' willingness to pay more for the brand-name good, these critics assert, is a form of irrationality fostered by advertising. Economist Edward Chamberlin, one of the early developers of the theory of monopolistic competition, concluded from this argument that brand names were bad for the economy. He proposed that the government discourage their use by refusing to enforce the exclusive trademarks that companies use to identify their products.

More recently, economists have defended brand names as a useful way for consumers to ensure that the goods they buy are of high quality. There are two related arguments. First, brand names provide consumers with *information* which cannot be easily judged in advance of purchase. Second, brand names give firms an *incentive* to meet the needs of consumers, because firms have a financial stake in maintaining the reputation of their brand names. Note that branding does not always equate to high quality. Branding is primarily a means by which the firm creates an association in the consumer and because that association becomes familiar, the consumer is more likely to retain some loyalty to the firm and thus repeat purchase. Some firms, for example, will happily admit their goods are 'cheap and cheerful' but the key association in consumers' minds is one of value for money. Firms such as Lidl, Netto, Poundstretcher and Poundland are as much interested in developing their brand names as are Armani and Ralph Lauren.

To see how these arguments work in practice, consider a famous brand name: Ibis hotels. Imagine that you are driving through an unfamiliar town and you need somewhere to stay for the night. You see a Hotel

Ibis and a local hotel next door to it. Which do you choose? The local hotel may in fact offer better accommodation at lower prices, but you have no way of knowing that. In contrast, Hotel Ibis offers a consistent product across many European cities. Its brand name is useful to you as a way of judging the quality of what you are about to buy.

The Ibis brand name also ensures that the company has an incentive to maintain quality (In reality, hotel chains such as Ibis provide the opportunity for private firms to operate a hotel under the brand name and use the benefits of the brand name and be committed to meet its associated standards). For example, if some customers were to become very ill from bad food served at breakfast at an Ibis hotel, the news would be disastrous for the company. Ibis would lose much of the valuable reputation that it has built up over the years and, as a result, it would lose sales and profit not just in the hotel that served the bad food but in its many hotels across Europe. By contrast, if some customers were to become ill from bad food served at breakfast in a local hotel, that restaurant might have to close down, but the lost profits would be much smaller. Hence, Ibis has a greater incentive to ensure that its breakfast food is safe and that standards are maintained in all the hotels bearing its brand.

The debate over brand names thus centres on the question of whether consumers are rational in preferring brand names over generic substitutes. Critics argue that brand names are the result of an irrational consumer response to advertising. Defenders of brand names argue that consumers have good reason to pay more for brand-name products because they can be more confident in the quality of these products.

CONTESTABLE MARKETS

Most economics textbooks up to the late 1970s covered market structures ranging from perfect competition at one extreme to monopoly at the other. Changes in the way businesses actually operated in the real world meant that there were some gaps between the theory and the observed behaviour of firms. This led to the development of a new theory which was incorporated into the explanation of market structures. The theory of contestable markets was developed by William J. Baumol, John Panzar and Robert Willig in 1982.

The key characteristic of a perfectly contestable market (the benchmark to explain firms' behaviours) was that firms were influenced by the threat of new entrants into a market. The more highly contestable a market is, the lower the barriers to entry. We have seen how, in monopolistically competitive markets, despite the fact that each firm has some monopoly control over its product, the ease of entry and exit means that in the long run profits can be competed away as new firms enter the market. This threat of new entrants may make firms behave in a way that departs from what was assumed to be the traditional goal of firms – to maximize profits. The suggestion by Baumol and colleagues was that firms may deliberately limit profits made to discourage new entrants. Profits might be limited by what is termed **entry limit pricing**. This refers to a situation where a firm will keep prices lower than they could be in order to deter new entrants. Similarly, firms may also practise **predatory or destroyer pricing** whereby the price is held below average cost for a period to try and force out competitors or prevent new firms from entering the market. Incumbent firms may be in a position to do this because they may have been able to gain some advantages of economies of scale which new entrants may not be able to exploit.

entry limit pricing a situation where a firm will keep prices lower than they could be in order to deter new entrants

predatory or destroyer pricing a situation where firms hold price below average cost for a period to try and force out competitors or prevent new firms from entering the market

In a contestable market firms may also erect other artificial barriers to prevent entry into the industry by new firms. Such barriers might include operating at over-capacity, which provides the opportunity to flood the market and drive down price in the event of a threat of entry. Firms will also carry out aggressive marketing and branding strategies to ‘tighten’ up the market or find ways of reducing costs and increasing efficiency to gain competitive advantage. Searching out sources of competitive advantage is a topic written on extensively by Michael Porter, who defined **competitive advantage** as being the advantages firms can gain over another which are both distinctive and defensible. These sources are not simply to be found in terms of new product development but through close investigation and analysis of the supply chain, where little changes might make a difference to the cost base of a firm which it can then exploit to its advantage.

competitive advantage the advantages firms can gain over another which have the characteristics of being both distinctive and defensible

Hit-and-run tactics might be evident in a contestable market where firms enter the industry, take the profit and get out quickly (possible because of the freedom of entry and exit). In other cases firms may indulge in what is termed **cream-skimming** – identifying parts of the market that are high in value added and exploiting those markets.

cream-skimming a situation where a firm identifies parts of a market that are high in value added and seeks to exploit those markets

The theory of contestable markets has been widely adopted as a beneficial addition to the theory of the firm and there has been extensive research into its application. There are numerous examples of markets exhibiting contestability characteristics including financial services; airlines, especially flights on domestic routes; the IT industry and in particular internet service providers (ISPs), software and web developers; energy supplies and the postal service. The key to analysing market structures, therefore, could be argued to be the focus on the degree of freedom of entry and exit. If policymakers can keep barriers to entry as low as possible, i.e. try to ensure a high degree of contestability, then there is more likelihood that market outcomes will be efficient.

SELF TEST How might advertising make markets less competitive? How might it make markets more competitive? • Give the arguments for and against brand names.

CONCLUSION

Monopolistic competition is true to its name: it is a hybrid of monopoly and competition. Like a monopoly, each monopolistic competitor faces a downwards sloping demand curve and, as a result, charges a price above marginal cost. As in a perfectly competitive market, there are many firms, and entry and exit drive the profit of each monopolistic competitor toward zero. Table 15.2 summarizes these lessons.

Because monopolistically competitive firms produce differentiated products, each firm advertises in order to attract customers to its own brand. To some extent, advertising manipulates consumers’ tastes, promotes irrational brand loyalty and impedes competition. Equally, advertising can provide information, establish brand names of reliable quality and foster competition.

The theory of monopolistic competition seems to describe many markets in the economy. It is somewhat disappointing, therefore, that the theory does not yield simple and compelling advice for public policy. From the standpoint of the economic theorist, the allocation of resources in monopolistically competitive markets is not perfect. Yet, from the standpoint of a practical policymaker, there may be little that can be done to improve it.

TABLE 15.2

Monopolistic Competition: Between Perfect Competition and Monopoly Market structure

| | Market structure | | |
|---|---------------------|--------------------------|------------------|
| | Perfect competition | Monopolistic competition | Monopoly |
| Features that all three market structures share | | | |
| Goal of firms | Maximize profits | Maximize profits | Maximize profits |
| Rule for maximizing | $MR = MC$ | $MR = MC$ | $MR = MC$ |
| Can earn economic profits in the short run? | Yes | Yes | Yes |
| Features that monopoly and monopolistic competition share | | | |
| Price taker? | Yes | No | No |
| Price | $P = MC$ | $P > MC$ | $P > MC$ |
| Produces welfare-maximizing level of output? | Yes | No | No |
| Features that perfect competition and monopolistic competition share | | | |
| Number of firms | Many | Many | One |
| Entry in long run? | Yes | Yes | No |
| Can earn economic profits in long run? | No | No | Yes |

IN THE NEWS

Branding

A key element of branding is to differentiate yourself from rivals and also to find a way of generating some customer loyalty. There are plenty of different ways to do this, some not as obvious as others.

Best Brands are Disruptively Consistent

I'll never write a book on branding. I don't have the time to sit down and commit the months, probably years, it would take to create something of value on the topic of brand management. And the people I have spoken to who have committed that kind of time to writing a management book invariably regret it down the track.

But if I did write a book on branding I would call it *Disruptively Consistent*. And if it sold well I would bring out a sequel called *Consistently Disruptive*. The concept is the same whichever way you write it and I fundamentally believe that this

paradoxical concept is at the very centre of all great brand strategy.

Let me explain with one of my favourite case studies. Ten years ago, office retailer Staples had a problem. Its original positioning of offering the widest ranges at the lowest prices had worked so well against small, independent competitors that it had almost wiped them out. The problem for Staples was that this same positioning was now failing to differentiate it against its two remaining large national rivals – Office World and Office Depot. Worse still, the pricing part of that promise was now leading to some less than attractive margins as it went head to head with these two big competitors.

When Shira Goodman took over as Staples chief marketing officer (CMO), she quickly realized she had to come up with a new position for the brand given the current approach was no longer working. After months of research, the decision was made to focus away from price and range, given that they had become points of parity, and focus instead on an association that none of the big three players currently owned – easy. Staples would make buying office products easy.

Staples set about this new positioning challenge with vigour. It created a new logo with the strap line 'that was easy' under it. It created the now famous Easy Button as a symbol. It launched a series of ads in which customers were shown being

(Continued)

rescued from the complicated hell of office products by Staples and its Easy Button. It even launched an Invention Quest competition in which a prize was awarded each year for an entrepreneur who invented an office product that would make life easier for consumers. All bang on the brand.

But it was the smallest and most subtle change that made the most impact. When Goodman reviewed the existing design of the Staples stores, she noticed that all the most popular products were right at the back of the store. When she inquired further she was told that this was standard retail practice. In a supermarket, for example, milk is always at the back of the store because it ensures that consumers have to walk past aisles of crap they do not need, some of which they hopefully pop into their trolley. And it was the same with office retail. It was just the way that retailing was done.

But Goodman realized this was not easy. It was the exact opposite of that. And she began to move all the best-selling products to the front of the store to make the

Staples experience easier. Staples would lose a little in ancillary sales, but it would win the bigger and more important battle of brand differentiation.

And it worked. Ten years on, Staples stands as the undisputed market leader in office products. Staples' regular brand tracking now reveals the brand owns the association of 'easy' and that has driven enormous market share gains.

And if you review why Staples has been so successful in differentiating itself, it all comes back to little decisions like redesigning its stores so that the best-selling products are the easiest to reach. And we return to the title of the book that I will never write – *Disruptively Consistent*.

Brands win when they disrupt the category they operate in. One of the sure signs that a brand is truly strong is when it defies the traditional logic of its competitors and breaks some of the central rules of the industry it is in.

But this rule breaking is not arbitrary. Not chaotic. The only reason a strong brand breaks the rules of the industry is to deliver their brand

positioning to their target consumers. At the heart of any strong brand, there are two motions taking place simultaneously. On the one hand, the brand is disrupting its category. On the other, it is ruthlessly executing its brand positioning. And the place where these two motions meet is where great brands exist.

I could give you many more examples. The champagne brand that reversed the sequence of its vintages, in direct contradiction of the rules of its category. Or the airline that replaced its boring arrival announcements with a more cheerful, entertaining message because 'fun' was one of its five core values. All these, and more, are in the book that I will never write.

The central message of branding remains the challenge of doing two things well. First, ensuring that the positioning is applied steadfastly and without exception. Second, breaking as many rules as possible within the business to achieve that end. Or disruptively consistent as my never-to-be written book would put it.

Questions

- 1 In what way can branding be 'disruptive' in a monopolistically competitive market?
- 2 Why do you think that Staples, strategy of offering low prices and a wide range was successful against smaller firms in the industry?
- 3 Why do you think this strategy was unlikely to work against Staples, two larger rivals?
- 4 In what ways did the new CMO at Staples seek to differentiate the business from its rivals?
- 5 To what extent would you describe the office products market as highly contestable?

Staples Office Superstores – disruptively consistent.



SUMMARY

- A monopolistically competitive market is characterized by three attributes: many firms, differentiated products and free entry.
- The equilibrium in a monopolistically competitive market differs from that in a perfectly competitive market in two related ways. First, each firm in a monopolistically competitive market has excess capacity. That is, it operates on the downwards sloping portion of the average total cost curve. Second, each firm charges a price above marginal cost.
- Monopolistic competition does not have all the desirable properties of perfect competition. There is the standard deadweight loss of monopoly caused by the mark-up of price over marginal cost. In addition, the number of firms (and thus the variety of products) can be too large or too small. In practice, the ability of policymakers to correct these inefficiencies is limited.
- The product differentiation inherent in monopolistic competition leads to the use of advertising and brand names. Critics of advertising and brand names argue that firms use them to take advantage of consumer irrationality and to reduce competition. Defenders of advertising and brand names argue that firms use them to inform consumers and to compete more vigorously on price and product quality.
- The theory of contestable markets suggests that firm behaviour may be directed at pricing strategies designed to prevent new entrants or to exploit the value in markets. Public policy, therefore, might be more directed at focusing on keeping barriers to entry as low as possible so that prices are driven down towards average cost. Low barriers to entry provide a constraint on firms seeking to adopt strategies to prevent competition.

QUESTIONS FOR REVIEW

- 1 Describe the three attributes of monopolistic competition. How is monopolistic competition like monopoly? How is it like perfect competition?
- 2 Identify five ways a firm in a monopolistically competitive market might seek to differentiate its products from its rivals?
- 3 Draw a diagram depicting a firm in a monopolistically competitive market that is making profits. Now show what happens to this firm as new firms enter the industry.
- 4 Draw a diagram of the long-run equilibrium in a monopolistically competitive market. How is price related to average total cost? How is price related to marginal cost?
- 5 Does a monopolistic competitor produce too much or too little output compared to the most efficient level? What practical considerations make it difficult for policymakers to solve this problem?
- 6 Is the purpose of advertising simply to shift the demand curve for a product to the right?
- 7 How might advertising reduce economic well-being? How might advertising increase economic well-being?
- 8 How might advertising with no apparent informational content in fact convey information to consumers?
- 9 Explain two benefits that might arise from the existence of brand names.
- 10 In what ways does the theory of contestable markets suggest that firm behaviour might be affected by the degree of freedom of entry and exit into a market?

PROBLEMS AND APPLICATIONS

- 1 Classify the following markets as perfectly competitive, monopolistic or monopolistically competitive, and explain your answers:
 - a. wooden HB pencils
 - b. bottled water
 - c. copper
 - d. local telephone service
 - e. strawberry jam
 - f. lipstick.
- 2 What feature of the product being sold distinguishes a monopolistically competitive firm from a monopoly firm?
 - 3 The chapter states that monopolistically competitive firms could increase the quantity they produce and lower the average total cost of production. Why don't they do so?
 - 4 Sparkle is one firm of many in the market for toothpaste, which is in long-run equilibrium.
 - a. Draw a diagram showing Sparkle's demand curve, marginal revenue curve, average total cost curve and marginal cost curve. Label Sparkle's profit-maximizing output and price.
 - b. What is Sparkle's profit? Explain.

- c. On your diagram, show the consumer surplus derived from the purchase of Sparkle toothpaste. Also show the deadweight loss relative to the efficient level of output.
- d. If the government forced Sparkle to produce the efficient level of output, what would happen to the firm? What would happen to Sparkle's customers?
- 5 Do monopolistically competitive markets typically have the optimal number of products? Explain.
- 6 The chapter says that monopolistically competitive firms may send greetings cards to their customers. What do they accomplish by this? Explain in words and with a diagram.
- 7 If you were thinking of entering the ice-cream business, would you try to make ice cream that is just like one of the existing (successful) brands? Explain your decision using the ideas of this chapter.
- 8 Describe three adverts that you have seen on TV. In what ways, if any, were each of these adverts socially useful? In what ways were they socially wasteful? Did the adverts affect the likelihood of your buying the product? Why or why not?
- 9 For each of the following pairs of firms, explain which firm would be more likely to engage in advertising:
- A family-owned farm or a family-owned restaurant.
 - A manufacturer of forklift trucks or a manufacturer of cars.
 - A company that invented a very reliable watch or a company that invented a less reliable watch that costs the same amount to make.
- 10 The makers of *Panadol* pain reliever do a lot of advertising and have very loyal customers. In contrast, the makers of generic paracetamol do no advertising, and their customers shop only for the lowest price. Assume that the marginal costs of Panadol and generic paracetamol are the same and constant.
- Draw a diagram showing *Panadol's* demand, marginal revenue and marginal cost curves. Label *Panadol's* price and mark-up over marginal cost.
 - Repeat part (a) for a producer of generic paracetamol. How do the diagrams differ? Which company has the bigger mark-up? Explain.
 - Which company has the bigger incentive for careful quality control? Why?
 - How might barriers to entry influence the behaviour of the makers of *Panadol*?
 - What factors would affect the extent to which the makers of *Panadol* could engage in predatory or destroyer pricing to force out competitors in this market.

16

MARKET STRUCTURES III: OLIGOPOLY

The Europeans love chocolate. The average German eats about 180 62-gram bars of chocolate a year. The Belgians are not far behind at 177 bars, the Swiss around 173 and the British eat around 164 bars per year. There are many firms producing chocolate in Europe including Anthon Berg in Denmark, Camille Bloch, Lindt and Favarger in Switzerland, Guylian and Godiva in Belgium, and Hachez in Germany. However, Europeans are liable to find that what they are eating is likely to be made by one of three companies: Cadbury (now owned by the U.S firm Kraft), Mars or Nestlé. These firms dominate the chocolate industry in the European Union. Being so large and dominant they are able to influence the quantity of chocolate bars produced and, given the market demand curve, the price at which chocolate bars are sold.

The European market for chocolate bars is a further example of imperfect competition but in this case the market is dominated by a relatively small number of very large firms. This type of imperfect competition is referred to as **oligopoly** – competition amongst the few. In oligopolistic markets, there might be many thousands of firms in the industry but sales are dominated by a small number of firms. The market is said to be concentrated in the hands of a few firms. The **concentration ratio** refers to the proportion of total market share accounted for by a particular number of firms. A two-firm concentration ratio of 90 per cent, for example, means that 90 per cent of all sales in the market are accounted for by just two firms. A five-firm concentration ratio of 75 per cent means that three-quarters of all sales in the market are accounted for by five firms. The small number of dominant sellers makes rigorous competition less likely, and it makes strategic interactions among them vitally important. As a result, the actions of any one seller in the market can have a large impact on the profits of all the other sellers. That is, oligopolistic firms are interdependent in a way that competitive firms are not. Our goal in this chapter is to see how this interdependence shapes the firms' behaviour and what problems it raises for public policy.

oligopoly competition amongst the few – a market structure in which only a few sellers offer similar or identical products and dominate the market

concentration ratio the proportion of total market share accounted for by a particular number of firms

CHARACTERISTICS OF OLIGOPOLY

The main characteristic of oligopolistic markets is that there are a relatively small number of dominant firms in the market. Each firm may offer a product similar or identical to the others. One example is the market for chocolate bars. Other examples include the world market for crude oil – a few countries in the Middle East control much of the world's oil reserves – and supermarkets in parts of Europe. In the UK, for example, there are many thousands of firms selling groceries but the industry is dominated by four very large firms: Tesco, Sainsbury's, Morrisons and Asda. There are approximately a dozen companies that now sell cars in Europe so whether this can be described as an oligopoly is open to debate. There are thousands of small independent breweries across Europe but sales are dominated by a relatively small number of firms, and are dominated by A-BInBev, Heineken, Carlsberg and SABMiller.

SELF TEST Think of the following different markets which can be classed as being oligopolistic market structures and what the approximate concentration ratios are: banking, mobile phone networks, insurance, the chemical industry, electrical goods, detergents, and entertainment.

Differentiation

Firms in oligopolistic market structures do sell products that are similar but may seek to differentiate themselves in some way. One lager, for example, is not that dissimilar to many others but somehow brewing firms have to try and convince customers that their particular lager is different. This may be done through making the alcohol content higher or lower (or even zero alcohol), having a beer that is light in calories, in the way the product is packaged, how the beer stays fizzy and so on.

Firms such as Procter & Gamble produce household care products including Daz, Ariel, Bold and Fairy washing products. These are essentially all products to wash clothes but Procter & Gamble find ways to differentiate this product portfolio and even within a particular brand produce different versions of each brand such as washing powders, tablets and liquids, with or without fabric softener and/or stain remover. Procter & Gamble are trying to differentiate its products to meet different customer needs and market segments and also to capture market share from its rivals. **Market segments** refer to the way in which firms break down customers into groups with similar buying habits or characteristics such as age, culture, gender, income, location, aspiration, interest, background and so on.

market segments the breaking down of customers into groups with similar buying habits or characteristics

Interdependence

Because oligopolistic markets are dominated by a few large firms, they are said to be interdependent. This means that what one firm does has some influence on the others and each firm may or may not react to the decisions of others. Each firm in the industry will be considering its own actions but its behaviour will be influenced by what it thinks the reaction of its rivals will be.

A result of this interdependence is that tensions can arise between firms of cooperation and self-interest. The group of oligopolists is best off cooperating and acting like a monopolist – producing a small quantity of output and charging a price above marginal cost. Yet because each oligopolist cares about only its own profit, there are powerful incentives at work that hinder a group of firms from maintaining the monopoly outcome.

A Duopoly Example

To understand the behaviour of oligopolies, let's consider an oligopoly with only two members, called a *duopoly*. Duopoly is the simplest type of oligopoly. Oligopolies with three or more members face the same problems as oligopolies with only two members, so we do not lose much by starting with the case of duopoly.

Imagine a town in which only two residents – Jacques and Joelle – own wells that produce water safe for drinking. Each Saturday, Jacques and Joelle decide how many litres of water to pump, bring the water to town, and sell it for whatever price the market will bear. To keep things simple, suppose that Jacques and Joelle can pump as much water as they want without cost. That is, the marginal cost of water equals zero.

Table 16.1 shows the town's demand schedule for water. The first column shows the total quantity demanded, and the second column shows the price. If the two well owners sell a total of 10 litres of water, water sells for €110 a litre. If they sell a total of 20 litres, the price falls to €100 a litre. And so on. If you graphed these two columns of numbers, you would get a standard downwards sloping demand curve.

The last column in Table 16.1 shows the total revenue from the sale of water. It equals the quantity sold times the price. Because there is no cost to pumping water, the total revenue of the two producers equals their total profit.

TABLE 16.1 The Demand Schedule for Water

| Quantity (in litres) | Price (€) | Total revenue (and total profit €) |
|-------------------------|--------------|---------------------------------------|
| 0 | 120 | 0 |
| 10 | 110 | 1,100 |
| 20 | 100 | 2,000 |
| 30 | 90 | 2,700 |
| 40 | 80 | 3,200 |
| 50 | 70 | 3,500 |
| 60 | 60 | 3,600 |
| 70 | 50 | 3,500 |
| 80 | 40 | 3,200 |
| 90 | 30 | 2,700 |
| 100 | 20 | 2,000 |
| 110 | 10 | 1,100 |
| 120 | 0 | 0 |

Let's now consider how the organization of the town's water industry affects the price of water and the quantity of water sold.

Competition, Monopolies and Cartels

In a competitive market, the production decisions of each firm drive price equal to marginal cost. In the market for water, marginal cost is zero. Thus, under competition, the equilibrium price of water would be zero, and the equilibrium quantity would be 120 litres. The price of water would reflect the cost of producing it, and the efficient quantity of water would be produced and consumed.

Table 16.1 shows that total profit is maximized at a quantity of 60 litres and a price of €60 a litre. A profit-maximizing monopolist, therefore, would produce this quantity and charge this price. As is standard for monopolies, price would exceed marginal cost. The result would be inefficient, for the quantity of water produced and consumed would fall short of the socially efficient level of 120 litres.

What outcome should we expect from our duopolists? The tension between self-interest and cooperation exists because of the characteristic of interdependence and so one possibility is that Jacques and Joelle get together and agree on the quantity of water to produce and the price to charge for it. Such an agreement among firms over production and price is called **collusion**, and the group of firms acting in unison is called a **cartel**. Once a cartel is formed, the market is in effect served by a monopoly, and we can apply analysis assuming monopoly. That is, if Jacques and Joelle were to collude, they would agree on the monopoly outcome because that outcome maximizes the total profit that the producers can get from the market. Our two producers would produce a total of 60 litres, which would be sold at a price of €60 a litre. Once again, price exceeds marginal cost, and the outcome is socially inefficient.

collusion an agreement among firms in a market about quantities to produce or prices to charge
cartel a group of firms acting in unison

A cartel must agree not only on the total level of production but also on the amount produced by each member. In our case, Jacques and Joelle must agree how to split between themselves the monopoly production of 60 litres. Each member of the cartel will want a larger share of the market because a larger market share means larger profit. If Jacques and Joelle agreed to split the market equally, each would produce 30 litres, the price would be €60 a litre and each would get a profit of €1,800.

CASE STUDY China and LCD Cartels

Competition authorities around the world have investigated price fixing by manufacturers of liquid crystal display panels and, in 2010, the European Union competition commission fined six firms almost €650 million for being in a cartel. The firms, Samsung, LG, AU Optronics, Chimei InnoLux, Chungwha Picture Tubes and HannStar Display were accused of operating a cartel between 2001 and 2006. At the time, Samsung was given immunity from the fine because it provided evidence to the commission on the price fixing.

U.S regulators have also taken action against the manufacturers of these flat screen TVs and the total fines handed out across the world have been around €2.9 billion.

To this figure can now be added €43.5 million following the decision by China's economic planning agency, The National Development and Reform Commission, to fine six firms from South Korea and Taiwan for price fixing. The action by the Chinese authorities was welcomed across the world as an important step in signalling to firms that no part of the world would be prepared to tolerate anti-competitive behaviour. Critics at the time argued that the fine levied by the Chinese government was too small but the fact that the action was taken was seen as being more important.

If firms know that China's authorities will act against global cartels then this sends a powerful message. Typically, cartel arrangements are subject to investigation when one firm in the cartel goes to the authorities and acts as a 'whistle blower' in exchange for more lenient treatment. Across other parts of the world, this incentive for firms to renege exists and there is clear evidence that whistle blowers have been treated leniently as was the case for Samsung in Europe. No one really knew if the same principle would be applied by the Chinese authorities but now there is every chance that firms will think very carefully about getting involved in cartel activity for fear of who might renege on the agreement.



Go into any electronics retailer and you might observe that prices of goods are very similar even though the products are produced by different firms. Is this because of competition or are prices being fixed?

The Equilibrium for an Oligopoly

Although oligopolists would like to form cartels and earn monopoly profits, often that is not possible. Competition laws prohibit explicit agreements among oligopolists as a matter of public policy. In addition, squabbling among cartel members over how to divide the profit in the market sometimes makes agreement among them impossible. Let's therefore consider what happens if Jacques and Joelle decide separately how much water to produce.

At first one might expect Jacques and Joelle to reach the monopoly outcome on their own, for this outcome maximizes their joint profit. In the absence of a binding agreement, however, the monopoly outcome is unlikely. To see why, imagine that Jacques expects Joelle to produce only 30 litres (half of the monopoly quantity). Jacques would reason as follows:

I could produce 30 litres as well. In this case, a total of 60 litres of water would be sold at a price of €60 a litre. My profit would be €1,800 (30 litres × €60 a litre). Alternatively, I could produce 40 litres. In this case, a total of 70 litres of water would be sold at a price of €50 a litre. My profit would be €2,000

(40 litres × €50 a litre). Even though total profit in the market would fall, my profit would be higher, because I would have a larger share of the market.

Because of the interdependence between the two firms, Joelle might reason the same way. If so, Jacques and Joelle would each bring 40 litres to town. Total sales would be 80 litres, and the price would fall to €40. Thus, if the duopolists individually pursue their own self-interest when deciding how much to produce, they produce a total quantity greater than the monopoly quantity, charge a price lower than the monopoly price and earn total profit less than the monopoly profit.

Although the logic of self-interest increases the duopoly's output above the monopoly level, it does not push the duopolists to reach the competitive allocation. Consider what happens when each duopolist is producing 40 litres. The price is €40, and each duopolist makes a profit of €1,600. In this case, Jacques's self-interested logic leads to a different conclusion:

My profit is €1,600. Suppose I increase my production to 50 litres. In this case, a total of 90 litres of water would be sold, and the price would be €30 a litre. Then my profit would be only €1,500. Rather than increasing production and driving down the price, I am better off keeping my production at 40 litres.

The outcome in which Jacques and Joelle each produce 40 litres looks like some sort of equilibrium. In fact, this outcome is called a **Nash equilibrium** (named after economic theorist John Nash, whose life was portrayed in the book, *A Beautiful Mind*, and the film of the same name). A Nash equilibrium is a situation in which economic actors interacting with one another each choose their best strategy given the strategies the others have chosen. In this case, given that Joelle is producing 40 litres, the best strategy for Jacques is to produce 40 litres. Similarly, given that Jacques is producing 40 litres, the best strategy for Joelle is to produce 40 litres. Once they reach this Nash equilibrium, neither Jacques nor Joelle has an incentive to make a different decision.

Nash equilibrium a situation in which economic actors interacting with one another each choose their best strategy given the strategies that all the other actors have chosen

Oligopolists would be better off cooperating and reaching the monopoly outcome. Yet because they pursue their own self-interest, they do not end up reaching the monopoly outcome and maximizing their joint profit. Each oligopolist is tempted to raise production and capture a larger share of the market. As each of them tries to do this, total production rises and the price falls.

At the same time, self-interest does not drive the market all the way to the competitive outcome. Like monopolists, oligopolists are aware that increases in the amount they produce reduce the price of their product. Therefore, they stop short of following the competitive firm's rule of producing up to the point where price equals marginal cost.

In summary, when firms in an oligopoly individually choose production to maximize profit, they produce a quantity of output greater than the level produced by monopoly and less than the level produced by competition. The oligopoly price is less than the monopoly price but greater than the competitive price (which equals marginal cost).

How the Size of an Oligopoly Affects the Market Outcome

We can use the insights from this analysis of duopoly to discuss how the size of an oligopoly is likely to affect the outcome in a market. Suppose, for instance, that Monika and Liesel suddenly discover water sources on their property and join Jacques and Joelle in the water oligopoly. The demand schedule in Table 16.1 remains the same, but now more producers are available to satisfy this demand. How would an increase in the number of sellers from two to four affect the price and quantity of water in the town?

If the sellers of water could form a cartel, they would once again try to maximize total profit by producing the monopoly quantity and charging the monopoly price. Just as when there were only two sellers, the members of the cartel would need to agree on production levels for each member and find some way to enforce the agreement. As the cartel grows larger, however, this outcome is less likely. Reaching and enforcing an agreement becomes more difficult as the size of the group increases.

If the oligopolists do not form a cartel – perhaps because competition laws prohibit it – they must each decide on their own how much water to produce. To see how the increase in the number of sellers affects

the outcome, consider the decision facing each seller. At any time, each well owner has the option to raise production by 1 litre. In making this decision, the well owner weighs two effects:

- *The output effect.* Because price is above marginal cost, selling one more litre of water at the going price will raise profit.
- *The price effect.* Raising production will increase the total amount sold, which will lower the price of water and lower the profit on all the other litres sold.

If the output effect is larger than the price effect, the well owner will increase production. If the price effect is larger than the output effect, the owner will not raise production. (In fact, in this case, it is profitable to reduce production.) Each oligopolist continues to increase production until these two marginal effects exactly balance, taking the other firms' production as given.

Now consider how the number of firms in the industry affects the marginal analysis of each oligopolist. The larger the number of sellers, the less concerned each seller is about its own impact on the market price. That is, as the oligopoly grows in size, the magnitude of the price effect falls. When the oligopoly grows very large, the price effect disappears altogether, leaving only the output effect. In this extreme case, each firm in the oligopoly increases production as long as price is above marginal cost.

We can now see that a large oligopoly is essentially a group of competitive firms. A competitive firm considers only the output effect when deciding how much to produce: because a competitive firm is a price taker, the price effect is absent. Thus, as the number of sellers in an oligopoly grows larger, an oligopolistic market looks more and more like a competitive market. The price approaches marginal cost and the quantity produced approaches the socially efficient level.

The Effects of International Trade Imagine that Toyota and Honda are the only car manufacturers in Japan, Volkswagen and BMW are the only car manufacturers in Germany, and Citroën and Peugeot are the only car manufacturers in France. If these nations prohibited international trade in cars, each would have a motorcar oligopoly with only two members, and the market outcome would likely depart substantially from the competitive ideal. With international trade, however, the car market is a world market and the oligopoly in this example has six members. Allowing free trade increases the number of producers from which each consumer can choose, and this increased competition keeps prices closer to marginal cost. Thus, the theory of oligopoly provides a reason why all countries can benefit from free trade.

SELF TEST If the members of an oligopoly could agree on a total quantity to produce, what quantity would they choose? ● If the oligopolists do not act together but instead make production decisions individually, do they produce a total quantity more or less than in your answer to the previous question? Why?

Oligopolies and Kinked Demand Curves

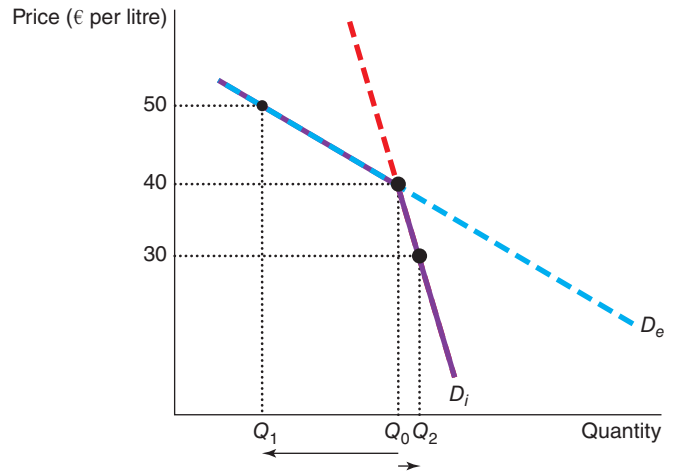
Interdependence can have other consequences which can be seen in the existence of so-called price rigidities or price stickiness in oligopolistic markets. This is the observations that in oligopolistic market structures prices do not change greatly over time and, given the changes in demand and supply in such markets, over time this might seem surprising. Here is a logic to the idea of price rigidities if we assume particular conditions; these are that the products being produced are homogenous and that if a firm raises its price its rivals will not follow suit but if a firm lowers its price its rivals will follow suit and all reduce their prices.

Assume that our duopolists, Jacques and Joelle, are currently charging a price of €40 per litre. Jacques is thinking about increasing price to €50 per litre but knows that if he does Joelle will keep her price at €40 per litre. In that situation, because the product is homogenous (i.e. undifferentiated) Jacques risks customers switching to Joelle and losing sales. The demand curve facing Jacques, is therefore, highly elastic at a price higher than €40 per litre.

Alternatively, Jacques might consider a strategy of reducing price from €40 to try and capture market share from Joelle. However, he knows that if he does reduce price to (say) €30 per litre, Joelle will follow suit and very quickly he will not gain any benefit from the lower price. Jacques effectively faces a highly inelastic demand curve at a price below €40. Our duopolists effectively face two demand curves as shown in Figure 16.1: one demand curve which is relatively elastic, D_e , and another which is relatively inelastic, D_i . If Jacques increases his price to €50, Joelle does not change her price and demand falls from Q_0 to Q_1 .

FIGURE 16.1**The Oligopolists Kinked Demand Curve**

Jacques faces two demand curves with different average elasticities. Demand curve D_e is relatively elastic and demand curve D_i is relatively inelastic. If Jacques increases or decreases price the response of Joelle means that his total revenue would fall in each case. There is no incentive for Jacques or Joelle to change price and the duopolists face a kinked demand curve.



The proportionate change in demand is greater than the increase in price and Jacques sees his total revenue fall. Equally, if Jacques reduces his price to €30, Joelle follows suit and so demand only rises by a small proportion to Q_2 from Q_0 . The proportionate increase in demand is less than the reduction in price and, again, Jacques sees his revenue fall.

If we assume that Jacques and Joelle both behave in the same way then there is no incentive for them to change price. Neither will want to put price up if they think the other will not follow suit and if either decides to reduce price the other will and so neither will benefit. The duopolists effectively face a kinked demand curve. At any price above €40 they face the elastic part of the demand curve D_e and at any price below €40 they face the demand curve D_i .

If the assumptions we have made above hold, then the focus of competitive behaviour of firms in an oligopoly might be on finding ways to differentiate their products rather than attempting to compete on price – so called **non-price competition**. Types of non-price competition may include a focus on quality, service, delivery, style, design, location, convenience, packaging, advertising and promotion, branding and so on.

non-price competition a situation where two or more firms seek to increase demand and market share by methods other than through changing price

There are limitations to the kinked demand curve theory which call into question the assumptions that underpin it. Firms may engage in collusion to set price or output, there is some evidence that if one firm increases its price others may follow suit, and how the firm finds the exact point of the kink in the demand curve is not clear.

GAME THEORY AND THE ECONOMICS OF COOPERATION

As we have seen, oligopolies would like to reach the monopoly outcome, but doing so requires cooperation, which at times is difficult to maintain. In this section we look more closely at the problems people face when cooperation is desirable but difficult. To analyse the economics of cooperation, we need to learn a little about game theory.

Game theory is the study of how people behave in strategic situations. By ‘strategic’ we mean a situation in which each person, when deciding what actions to take, must consider how others might respond to that action. Because the number of firms in an oligopolistic market is small, each firm must act strategically. Each firm knows that its profit depends not only on how much it produces but also on how much the other firms produce. In making its production decision, each firm in an oligopoly should consider how its decision might affect the production decisions of all the other firms.

game theory the study of how people behave in strategic situations

Game theory is extremely useful for understanding the behaviour of oligopolies. In the following section we present some of the principles of game theory which have been applied to firms in oligopolistic markets. In any game there are players or actors (which might be firms) who face various options in decision making which are called strategies. In making a decision (choosing a strategy) there are outcomes or payoffs that arise as a result of the decision. Each player is assumed to know their own mind and to be able to identify the payoff of the strategy they choose. However, each player also knows that their opponent or rival also faces the same decisions and strategies and that these will also have associated payoffs. This has been referred to as the 'I think he thinks that I think that he thinks that I think ...' scenario. Each player, therefore, has to put themselves into the position of the other player/s before deciding on a strategy. The choices are represented as a **payoff matrix** which is a table showing the possible combination of outcomes (payoffs) depending on the strategy chosen by each player.

payoff matrix a table showing the possible combination of outcomes (payoffs) depending on the strategy chosen by each player

For example, look at the payoff matrix represented by Figure 16.2.

FIGURE 16.2

Payoff Matrix

The matrix shows two players, Firm X and Firm Y and the decisions they can take with regard to keeping or breaking their agreement. The triangles show the payoff associated with their respective decisions.

| | | Firm Y | |
|--------|-----------------|--------------------------------|-------------------------------|
| | | Keep agreement | Break agreement |
| Firm X | Keep agreement | Profit = €100 Profit = €100 | Profit = €200 Profit = €50 |
| | Break agreement | Profit = €50 Profit = €200 | Profit = €25 Profit = €25 |

Imagine that the two players are firms X and Y, who enter into an agreement to fix the price in a market. The payoff to each firm is the profit they make as a result of the agreement. Firm X is represented on the vertical plane of the matrix. It has two strategies – keep the agreement or break the agreement. The payoffs it faces are represented in the orange triangles of the matrix. Firm Y is represented on the horizontal plane of the matrix and it also faces the same strategies. Its payoffs are represented in the blue triangles of the matrix.

Assume that both Firm X and Firm Y make the decision to keep to the agreement. The payoff to both is given by looking at the top left-hand quadrant and the payoff is that both firms take a profit of €100. If Firm X keeps to the agreement but Firm Y opts to break the agreement then the payoffs are given by the top right-hand quadrant. In this case, Firm X will gain a profit of €50 but Firm Y gets a profit of €200. If Firm Y keeps the agreement but Firm X breaks the agreement the payoffs are given by the bottom left quadrant and in this case Firm X gains a profit of €200 and Firm Y of €50. If both firms break the agreement the outcome is given by the bottom right quadrant and both earn a profit of €25.

The Prisoners' Dilemma

A particularly important 'game' is called the **prisoners' dilemma**. This game provides insight into the difficulty of maintaining cooperation. Many times in life, people fail to cooperate with one another even when cooperation would make them all better off. An oligopoly is just one example. The story of the prisoners' dilemma contains a general lesson that applies to any group trying to maintain cooperation among its members.

prisoners' dilemma a particular 'game' between two captured prisoners that illustrates why cooperation is difficult to maintain even when it is mutually beneficial

The prisoners' dilemma is a story about two criminals who have been captured by the police. Let's call them Mr Green and Mr Blue. The police have enough evidence to convict Mr Green and Mr Blue of a crime, illegal distribution of cannabis, so that each would spend a year in jail given existing sentencing rules. The police also suspect that the two criminals have committed a jewellery robbery together in which a victim was badly injured, but they lack hard evidence to convict them of this major crime. The police question Mr Green and Mr Blue in separate rooms, and they offer each of them the following deal:

With the evidence we have on the drug possession we can lock you up for 1 year. If you confess to the jewellery robbery and implicate your partner, however, we'll give you immunity and you can go free. Your partner will get 20 years in jail. But if you both confess to the crime, we won't need your testimony and we can avoid the cost of a trial, so you will each get an intermediate sentence of 8 years.

The possible outcomes, 1 year in prison, go free, etc. are referred to as payoffs. If Mr Green and Mr Blue, heartless criminals that they are, care only about their own sentences, what would you expect them to do? Would they confess or remain silent? Figure 16.3 shows their choices.

Each prisoner has two strategies: confess or remain silent. The sentence each prisoner gets depends on the strategy he chooses and the strategy chosen by his partner in crime.

Consider first Mr Green's decision. He reasons as follows:

I don't know what Mr Blue is going to do. If he remains silent, my best strategy is to confess, since then I'll go free rather than spending a year in jail. If he confesses, my best strategy is still to confess, since then I'll spend 8 years in jail rather than 20. So, regardless of what Mr Blue does, I am better off confessing.

FIGURE 16.3

The Prisoners' Dilemma

In this game between two criminals suspected of committing a crime, the sentence that each receives depends both on his decision whether to confess or remain silent and on the decision made by the other.

| | | Mr Green's decision | |
|--------------------|---------------|---|---|
| | | Confess | Remain silent |
| Mr Blue's decision | Confess | Mr Green gets 8 years Mr Blue gets 8 years | Mr Green gets 20 years Mr Blue goes free |
| | Remain silent | Mr Green goes free Mr Blue gets 20 years | Mr Green gets 1 year Mr Blue gets 1 year |

In the language of game theory, a strategy is called a **dominant strategy** if it is the best strategy for a player to follow regardless of the strategies pursued by other players. In this case, confessing is a dominant strategy for Mr Green. He spends less time in jail if he confesses, regardless of whether Mr Blue confesses or remains silent.

dominant strategy a strategy that is best for a player in a game regardless of the strategies chosen by the other players

Now consider Mr Blue's decision. He faces exactly the same choices as Mr Green, and he reasons in much the same way. Regardless of what Mr Green does, Mr Blue can reduce his time in jail by confessing. In other words, confessing is also a dominant strategy for Mr Blue.

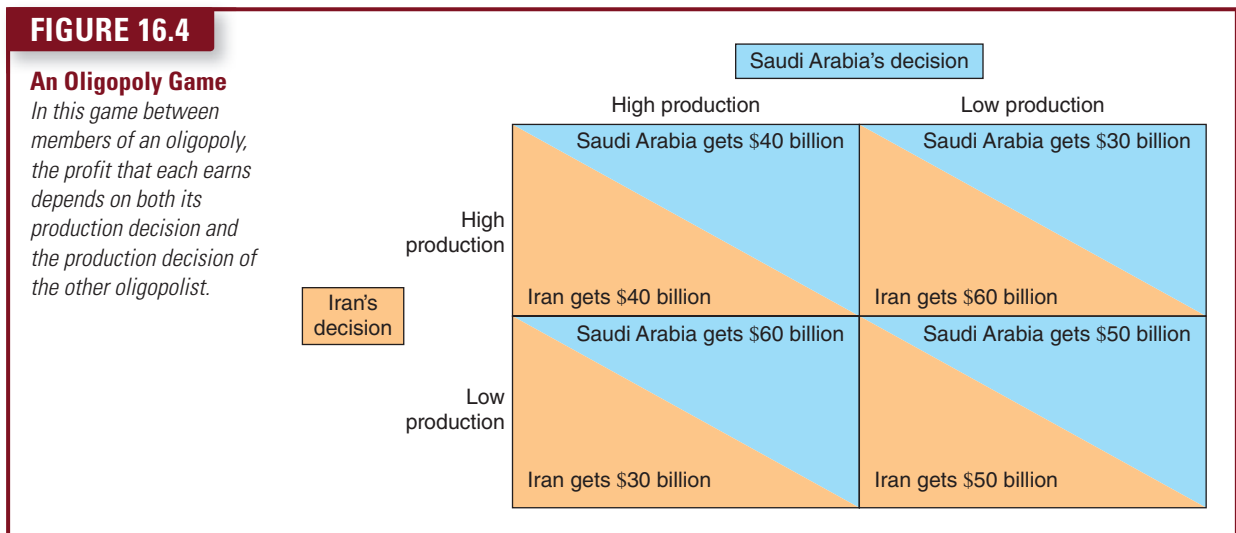
In the end, both Mr Green and Mr Blue confess, and both spend 8 years in jail. Yet, from their standpoint, this is a terrible outcome. If they had *both* remained silent, both of them would have been better off, spending only 1 year in jail on the drugs charge. By each pursuing his own interests, the two prisoners together reach an outcome that is worse for each of them.

To see how difficult it is to maintain cooperation, imagine that, before the police captured Mr Green and Mr Blue, the two criminals had made a pact not to confess. Clearly, this agreement would make them both better off *if* they both live up to it, because they would each spend only 1 year in jail. But would the two criminals in fact remain silent, simply because they had agreed to? Once they are being questioned separately, the logic of self-interest takes over and leads them to confess. Cooperation between the two prisoners is difficult to maintain, because cooperation is individually irrational.

Oligopolies as a Prisoners' Dilemma

The tension between self-interest and cooperation exemplified in the prisoners' dilemma is very similar to the tensions that exist between firms in imperfect competition and particularly between oligopolistic firms. Game theory has been applied extensively to the analysis of oligopolies as a result.

Consider an oligopoly with two countries: Iran and Saudi Arabia. Both countries sell crude oil. After prolonged negotiation, the countries agree to keep oil production low in order to keep the world price of oil high. After they agree on production levels, each country must decide whether to cooperate and live up to this agreement or to ignore it and produce at a higher level. Figure 16.4 shows the payoff matrix and how the profits of the two countries depend on the strategies they choose.



If both countries stick to their agreement they would earn \$50 billion in profit (the bottom right-hand quadrant). Suppose, however, you are the leader of Saudi Arabia. You might reason as follows:

I could keep production low as we agreed, or I could raise my production and sell more oil on world markets. If Iran lives up to the agreement and keeps its production low, then my country earns profit of \$60 billion (oil is priced in US dollars) with high production and \$50 billion with low production. In this case, my country is better off with high production.

If, however, Iran fails to live up to the agreement and produces at a high level, then my country earns \$40 billion with high production and \$30 billion with low production. Once again, my country is better off with high production. So, regardless of what Iran chooses to do, my country is better off renegeing on our agreement and producing at a high level is thus my dominant strategy.

Of course, Iran reasons in exactly the same way, and so both countries pursue their dominant strategy and produce at a high level. The result is the inferior outcome (from Iran and Saudi Arabia's standpoint) with each country earning \$40 billion in profits instead of the \$50 billion they could have earned if they had both stuck to their agreement.

This example illustrates why oligopolies have trouble maintaining monopoly profits. The monopoly outcome is jointly rational for the oligopoly, but each oligopolist has an incentive to cheat. Just as self-interest drives the prisoners in the prisoners' dilemma to confess, self-interest makes it difficult for the oligopoly to maintain the cooperative outcome with low production, high prices and monopoly profits.

Other Examples of the Prisoners' Dilemma

Advertising When two firms advertise to attract the same customers, they face a problem similar to the prisoners' dilemma. For example, consider the decisions facing two chemical companies: BASF and Evonik. If neither company advertises, the two companies split the market. If both advertise, they again split the market, but profits are lower, since each company must bear the cost of advertising. Yet if one company advertises while the other does not, the one that advertises attracts customers from the other. Figure 16.5 shows how the profits of the two companies depend on their actions. You can see that advertising is a dominant strategy for each firm. Thus, both firms choose to advertise, even though both firms would be better off if neither firm advertised.

FIGURE 16.5

An Advertising Game

In this game between firms selling similar products, the profit that each earns depends on both its own advertising decision and the advertising decision of the other firm.

| | | Evonik's decision | |
|-----------------|-----------------|--|--|
| | | Advertise | Don't advertise |
| BASF's decision | Advertise | BASF gets €3 million profit Evonik gets €3 million profit | BASF gets €5 million profit Evonik gets €2 million profit |
| | Don't advertise | BASF gets €2 million profit Evonik gets €5 million profit | BASF gets €4 million profit Evonik gets €4 million profit |

Common Resources We have seen that people tend to over-use common resources. One can view this problem as an example of the prisoners' dilemma. Imagine that two oil companies – Shell and BP – own adjacent oil fields. Under the fields is a common pool of oil worth €12 million. Drilling a well to recover the

oil costs €1 million. If each company drills one well, each will get half of the oil and earn a €5 million profit (€6 million in revenue minus €1 million in costs).

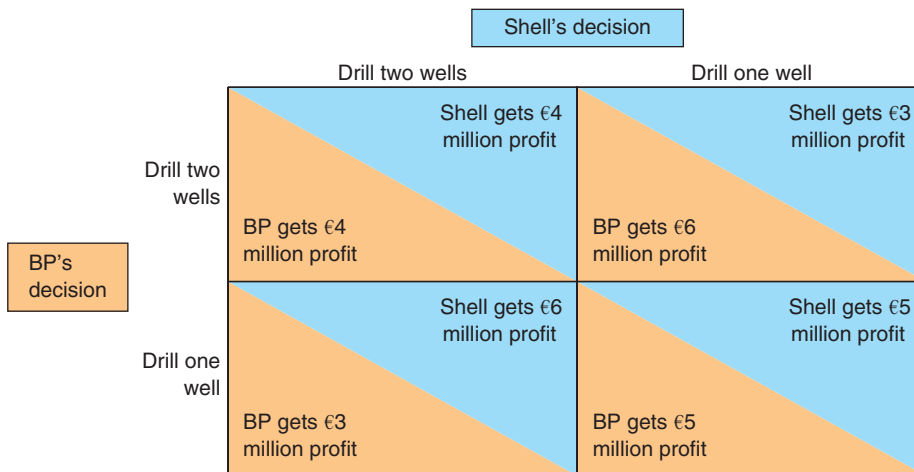
Because the pool of oil is a common resource, the companies will not use it efficiently. Suppose that either company could drill a second well. If one company has two of the three wells, that company gets two-thirds of the oil, which yields a profit of €6 million (two-thirds of €12 million = €8 million minus €2 million in costs to drill two wells). The other company gets the remaining one-third of the oil, for a profit of €3 million (€4 million minus €1 million for drilling the well). If each company drills a second well, the two companies again split the oil. In this case, each bears the cost of a second well, so profit is only €4 million for each company.

Figure 16.6 shows the game. Drilling two wells is a dominant strategy for each company. Once again, the self-interest of the two players leads them to an inferior outcome.

FIGURE 16.6

A Common Resources Game

In this game between firms pumping oil from a common pool, the profit that each earns depends on both the number of wells it drills and the number of wells drilled by the other firm.



Nash Equilibrium

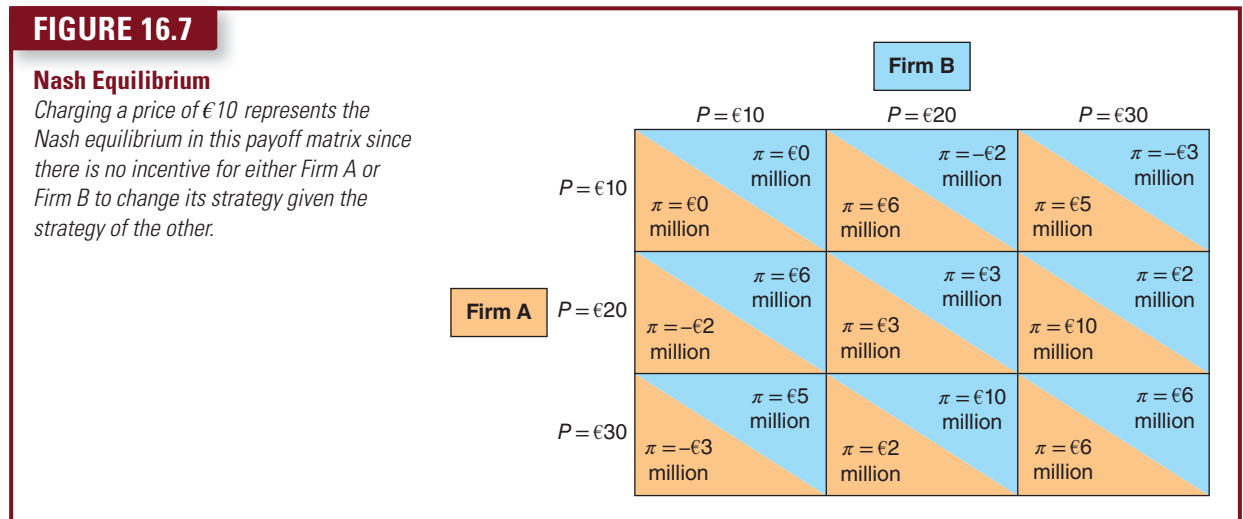
In our analysis of Jacques and Joelle (our economic actors), we mentioned that they reach an equilibrium which meant that neither has an incentive to choose any different strategy given the strategy that the other adopted. This was referred to as a Nash equilibrium after John Nash who developed the proposition.

The story of John Nash was dramatized in the film 'A Beautiful Mind' based on a book of the same name by Sylvia Nasar. Nash entered the world of game theory in the late 1940s and 1950s and gained the Nobel Prize for Economics in 1994 for his contribution to the field. (The scene in the film where Nash 'hits' upon his idea based on him and his friends choosing a girl to date is dramatic licence but an interesting way of presenting the theory all the same!) His work has had an impact on all manner of economic and political situations. It is important to remember that Nash was not an economist – he was a mathematician. In fact, Nash only took one course of economics during his undergraduate studies – a course in International Trade – and that was only to fulfil the degree requirements of the course he was on. Nash's ideas, therefore, are purely in the realm of the application of mathematical ideas to a bargaining problem. The application to economics and other areas has been somewhat broader.

At the heart of Nash's ideas was the mix of both cooperative and non-cooperative games. In the former, there are enforceable agreements between players (which may be in the form of legislation or the threat of fines from a regulator or similar) whilst in the latter there are not. The key thing in both cases is that the players in the game know that they cannot predict with any certainty what the other is going to do (exactly the situation facing firms in an oligopolistic market). Equally, they know what *they* want but are aware that all the other players think as they do.

The solutions that Nash derived were based around this thinking, where each player had to try to put him or herself in the position of others. The 'equilibrium' position would be where each player makes a decision which represents the best outcome in response to what other players' decisions are. The definition of a Nash equilibrium is a point where no player can improve their position by selecting any other available strategy whilst others are also playing their best option and not changing their strategies. One of the implications of Nash's work is that cooperation may well be the best option in the long term.

Let us take an example. Assume there are two firms competing with each other for profits in a market. The two firms have three decisions to make with regard to their pricing strategies. They can choose to set their price at either €10, €20 or €30. The payoff matrix showing the profits made at these different prices is shown in Figure 16.7.



If we look at the situation when each firm decides to set their price at €30, they both make a profit of €6 million (the bottom right-hand box). This, however, is not a Nash equilibrium, since Firm A could improve its position by reducing its price by €10 to €20, whilst A's strategy remains the same (setting price at €30). In this case, A would now gain a profit of €10 million rather than €6 million and B would be making €2 million as shown in the middle right box of the matrix.

Let's compare this to the situation where both firms set their price at €10. In this case, if Firm A decided to raise its price to €20, it would be worse off if B continued the strategy of charging €10. In such an instance, A would make a loss of €2 million whilst B would make €6 million. The same position applies to firm B; if it increased its price to €20 but A retained a price of €10, B would now make a loss of €2 million whilst A would make €6 million. There is no incentive, therefore, for either firm to change their position; the pay off of zero profit at a price of €10 represents a Nash equilibrium. Any other pricing decision by either firm would lead to one having an incentive to change its price to gain some advantage.

The price of €10, however, is not the best outcome for either firm. If they both agreed to charge a price of €30 they would both generate profits of €6 million.

Cooperative and Non-cooperative Games Two questions arise from this: can any such agreement be enforced and what happens if the game is repeated many times over? These were questions asked by

two other Nobel Prize winning economists, Thomas C. Schelling and Robert J. Aumann, who won the Prize in 2005. Schelling looked at non-cooperative and cooperative games. Cooperative game theory assumes that there is a set of outcomes or agreements that is known to each player and that each player has preferences over these outcomes. Non-cooperative game theory assumes players have a series of strategies they could use to gain an outcome and that each player has a preference over their desired outcome. The behaviour of firms or individuals might be affected by bargaining which entails some form of conflict of interest but in essence each player will be looking to maximize their returns, whilst knowing at the same time that some agreement is preferable to no agreement at all. In this scenario, how does a player manage to influence the negotiations in order to move towards his or her preferred outcome without upsetting the other players and thus failing to secure any agreement – an outcome which would be disadvantageous to all concerned, including the player?

Schelling proposed that it might be in the interests of the player to worsen his or her own options in order to gain some sort of concession from another player. Where difficulties arise is if both parties to a conflict make commitments that are seen as being irreversible and incompatible. The result could be stalemate and potential serious conflict.

In most 'game' situations, the protagonists know something about the position of the other – but not everything. However, if there is any perceived chink in the armour of the other player and this is detected by the other, then there is a potential benefit to follow the hard route. This is why this sort of game is referred to as 'chicken' or 'hawk/dove'. Schelling further included other complications to the analysis by looking at how the strategies of each player would change in light of threats and action. Schelling noted that parties will need to recognize that the costs to them of cheating or reneging and gaining some short-term benefit is far outweighed by the costs to them in the longer term of the destruction of the trust that results from cheating or reneging. The relationships between players will need to be assessed in the context of repeated playing of the game over a period of time.

This is the area that Aumann worked on – the field of long-term cooperation around game theory. We have already seen how, in the prisoners' dilemma, the dominant strategy for Mr Green and Mr Blue was to confess. Aumann asked the question about what the equilibrium outcome would be if the game were repeated over and over again, with each prisoner trying to maximize the *average* payoff from each game.

In such a situation, Aumann showed that the equilibrium outcome was to cooperate because any cheating on the agreement in the short term would be punishable by a refusal to cooperate at some point in the future – and both players would know this. Any short-term gains, therefore, are outweighed by longer-term losses. Aumann expressed this through what he referred to as a 'supergame' – that is, looking at the collection of repeated games as a whole game in itself. Aumann's work was extended to look at how groups of players might react in such situations. In an agreement between oligopolists, for example, there is always the tendency or incentive for one firm to break the agreement to seek to gain some advantage in the market. Aumann's work suggested that long-term cooperation could be 'enforced' by the many, against the few who might be seeking to defect.

The work was extended in subsequent research to try to take into account the strategies players might adopt in repeated games with incomplete information. This provides an incentive to players to hide, or seek to conceal, information from their rivals. Firms are very keen to keep their costs to themselves. If one player does manage to access information about their rivals and has some form of strategic advantage, therefore, what is the best way to utilize this knowledge? If this situation arose, would playing your hand to gain short-term benefit reveal that you did actually know more than you were letting on? For the player who does not have the information they would like, could they discover anything about the player's position by reviewing the strategies and decisions made by that player in the past?

Such scenarios are relevant to the world of financial markets where the issue of how to manage people who have access to privileged information, such as information about market moves, potential mergers or takeovers, announcements about key corporate decisions or product launches, business plans and strategies and so on, which they can use for personal (or corporate) gain.

Aumann and Schelling's work has a great many applications in everyday life. Trade wars, price wars, negotiations over budgets, wage negotiations, discussions about environmental issues, merger discussions, social policy, monetary policy, fiscal policy, the work of the Competition Commission, EU fisheries debates and oligopolies are just some of the major areas where their work can be applied.

The Prisoners' Dilemma and the Welfare of Society

The prisoners' dilemma describes many of life's situations, and it shows that cooperation can be difficult to maintain, even when cooperation would make both players in the game better off. Clearly, this lack of cooperation is a problem for those involved in these situations. But is lack of cooperation a problem from the standpoint of society as a whole? The answer depends on the circumstances.

In some cases, the non-cooperative equilibrium is bad for society as well as the players. In the common resources game in Figure 16.6, the extra wells dug by Shell and BP are pure waste. In both cases, society would be better off if the two players could reach the cooperative outcome.

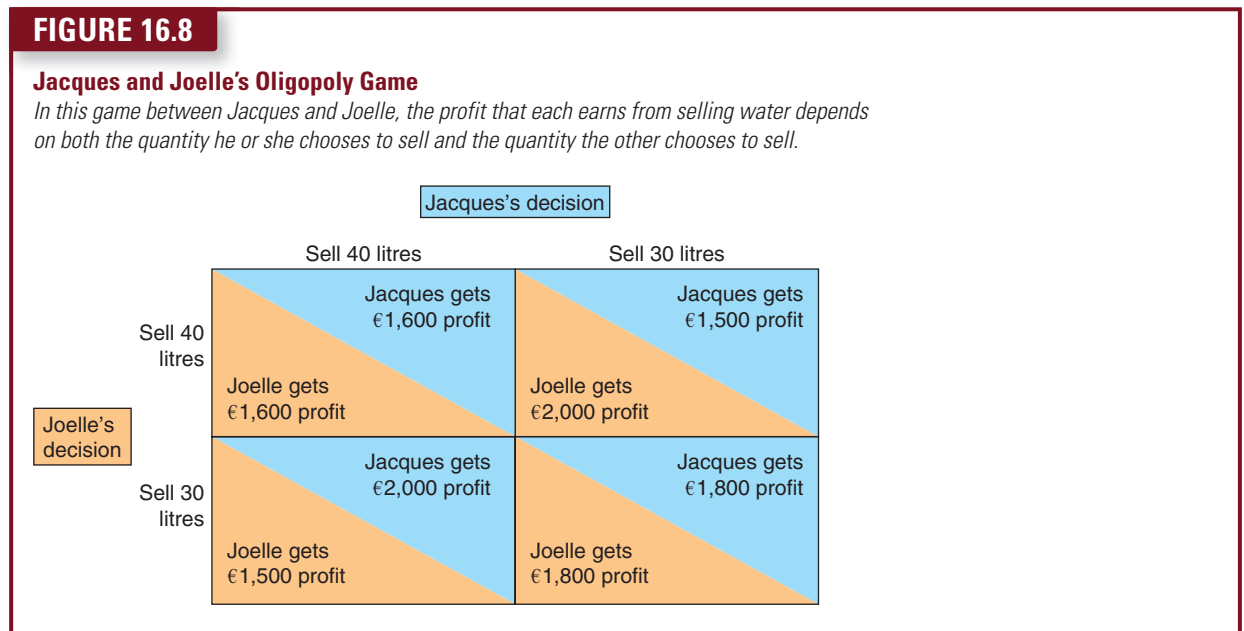
By contrast, in the case of oligopolists trying to maintain monopoly profits, lack of cooperation is desirable from the standpoint of society as a whole. The monopoly outcome is good for the oligopolists, but it is bad for the consumers of the product. The competitive outcome is best for society because it maximizes total surplus. When oligopolists fail to cooperate, the quantity they produce is closer to this optimal level. Put differently, the invisible hand guides markets to allocate resources efficiently only when markets are competitive, and markets are competitive only when firms in the market fail to cooperate with one another.

Similarly, consider the case of the police questioning two suspects. Lack of cooperation between the suspects is desirable, for it allows the police to convict more criminals. The prisoners' dilemma is a dilemma for the prisoners, but it can be a boon to everyone else.

Why People Sometimes Cooperate

The prisoners' dilemma shows that cooperation is difficult. But is it impossible? Not all prisoners, when questioned by the police, decide to turn in their partners in crime. Cartels sometimes do manage to maintain collusive arrangements, despite the incentive for individual members to defect. Very often, the reason that players can solve the prisoners' dilemma is that they play the game not once but many times.

To see why cooperation is easier to enforce in repeated games, let's return to our duopolists, Jacques and Joelle. Recall that Jacques and Joelle would like to maintain the monopoly outcome in which each produces 30 litres, but self-interest drives them to an equilibrium in which each produces 40 litres. Figure 16.8 shows the game they play. Producing 40 litres is a dominant strategy for each player in this game.



Imagine that Jacques and Joelle try to form a cartel. To maximize total profit they would agree to the cooperative outcome in which each produces 30 litres. Yet, if Jacques and Joelle are to play this game only once, neither has any incentive to live up to this agreement. Self-interest drives each of them to renege

and produce 40 litres. If Jacques takes the decision to renege on the agreement and produce 40 litres he stands to earn €2,000 in profit. Joelle thinks exactly the same way and so they both end up producing 40 litres and earning €1,600 in profit.

Now suppose that Jacques and Joelle know that they will play the same game every week. When they make their initial agreement to keep production low, they can also specify what happens if one party reneges. They might agree, for instance, that once one of them reneges and produces 40 litres, both of them will produce 40 litres forever after. This penalty is easy to enforce, for if one party is producing at a high level, the other has every reason to do the same.

The threat of this penalty may be all that is needed to maintain cooperation. Each person knows that defecting would raise his or her profit from €1,800 to €2,000. But this benefit would last for only one week. Thereafter, profit would fall to €1,600 and stay there. As long as the players care enough about future profits, they will choose to forgo the one-time gain from defection. Thus, in a game of repeated prisoners' dilemma, the two players may well be able to reach the cooperative outcome.

Tacit Collusion A repeated games scenario might also lead to a market outcome in which some form of collusion is suspected but in fact has arisen out of firms recognizing that they are interdependent. When firm behaviour results in a market outcome that appears to be anti-competitive, but has arisen because firms acknowledge that they are interdependent, this is referred to as **tacit collusion**. An example can be seen in typical out-of-town shopping malls where several firms have outlets all selling similar goods – carpets, electrical goods, furniture and so on. Shopping around these stores customers might have their suspicions aroused by the fact that regardless of the store, the prices are all very similar and in some cases identical. Even the promotional material promising to refund the difference if the customer can find the same good elsewhere cheaper looks to be an empty promise given the price similarity.

tacit collusion when firm behaviour results in a market outcome that appears to be anti-competitive but has arisen because firms acknowledge that they are interdependent

Is this an example of collusion? Have all the firms got together and fixed prices? Not necessarily. This may be a case of tacit collusion. A firm selling plasma TVs in one part of the shopping mall is likely to be aware that if it charges a price higher than its rivals that it will lose sales but equally it know that if it tries and compete aggressively on price its rivals are likely to follow suit. Each retailer thinks the same way and so prices tend to be very similar across the mall. In some cases, retailers will attempt to signal to their rivals that they will respond in kind by offering price guarantees such as the offer of refunds (which effectively means the store is saying to its rivals, 'if you charge a lower price we will match it') and confidently predicting that customers will not find the product at a lower price anywhere else (within reason of course). These sorts of messages might be thought of as being targeted at the customer but they can also be seen as being targeted just as much at rivals hence the suggestion that collusion is taking place albeit the collusion is tacit.

SELF TEST Tell the story of the prisoners' dilemma. Write down a table showing the prisoners' choices and explain what outcome is likely. • What does the prisoners' dilemma teach us about oligopolies?

MODELS OF OLIGOPOLY

Firms are in a position to set output and use the demand curve to determine price or can set price and use the demand curve to determine output. Which strategy is used will determine how firms behave. There are two basic models of oligopoly which reflect these differences, the Cournot model and the Bertrand model.

The Cournot Model

Earlier we noted that firms are interdependent. Most firms will have a reasonable idea of the size of the market in which they operate and what position they hold in that market. For ease of analysis, we will assume a duopoly and no barriers to entry in the industry and so the two firms are not unduly worried by new entrants. Assume that Firm A has conducted research which shows that the size of the market it operates in is 1,000,000 units or a value of €1 million a year. We will also assume that marginal cost is constant.

Figure 16.9 illustrates this situation. If Firm A were the only producer in the market it would produce where the MR curve cuts the MC curve and supply the whole 1 million units at Q_1 . However, Firm A knows that Firm B also operates in the market and supplies 20 per cent of the market. The demand curve it faces is thus not D_1 but D_2 , which is referred to as a residual demand curve. D_2 has an associated marginal revenue curve MR_2 . Firm A's profit maximizing output is now Q_2 where MR_2 cuts the MC curve. **Residual demand** is defined as the difference between the market demand curve and the amount supplied by other firms in the market. The residual demand curve depends on the output decision of the other firms in the market. Firm A might expend some time and effort trying to find out or at the very least estimate what this output level might be. In other words, Firm A needs to have some idea of the residual demand curve it faces or risk producing an output which would drive down the market price it faced. If, for example, it produced 1 million units and Firm B produced 200,000 units then there would be excess supply and the price it faces would fall. Knowing what your rival is planning to do is important in adopting the right strategy to ensure profit maximizing output.

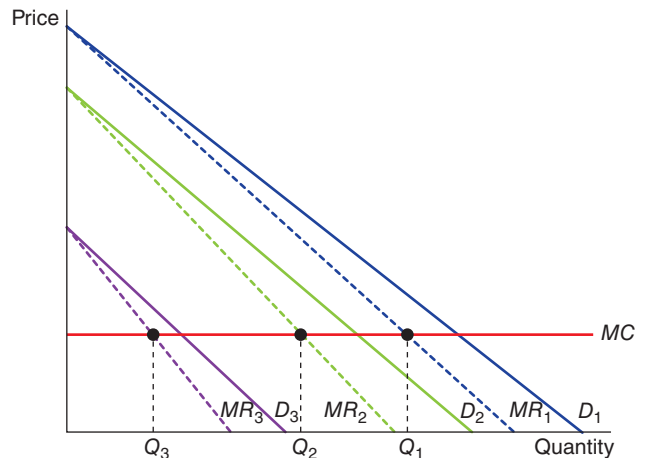
residual demand the difference between the market demand curve and the amount supplied by other firms in the market

If Firm B supplied 60 per cent of the market then Firm A would face a different residual demand curve D_3 , and would set output at Q_3 where MR_3 cuts the MC curve. It would be possible to conceptualize a situation where Firm A could map all possible outputs by Firm B and thus how it would react to these output levels.

FIGURE 16.9

Residual Demand

If Firm A supplied the whole market (i.e. it assumes Firm B produces nothing) then the profit maximizing output would be Q_1 . If Firm A assumes Firm B produces 20 per cent of the market (shown by demand curve D_2) then its new profit maximizing output will be Q_2 where the MR curve associated with the residual demand curve D_2 . If Firm B was assumed to produce 40 per cent of the market output then Firm A would produce where MR_3 associated with residual demand D_3 cuts the MC curve at output Q_3 .



The result of this analysis by Firm A would give it its reaction function. The **reaction function** shows the preferred strategies of one firm in relation to a particular issue which is dependent on the strategies adopted by its rival/s on that same issue. The reaction function, therefore, could outline the profit maximizing output for a firm given the simultaneous output decisions of its rivals.

reaction function the decision of one firm on a particular issue such as the profit maximizing output in response to the profit maximizing output decisions of its rivals

Firm B will also have a reaction function derived from its analysis of how it would react to the output decisions of Firm A. The respective reaction functions show how Firm A would react if Firm B changed its output decisions and vice versa. This model of oligopoly was developed by Augustin Cournot in 1838. Cournot assumed that given two firms in a duopoly, each firm determines its profit maximizing output on the assumption of the output of the other firm and that the decision of the other firm will not change in a given time period. The strategy each firm adopts, therefore, is based on setting the quantity of output. If we assume that both firms are profit maximizers then the question will arise as to what the profit maximizing output will be for Firm A given the output by Firm B?

In a given time period, Firm A, for example, could alter its output decision and its rival would not react. However, Firm B makes its decision under the same assumptions. This simultaneous decision making whereby each firm is trying to increase its profits but assuming its rivals will not react over different time periods eventually leads to an equilibrium position.

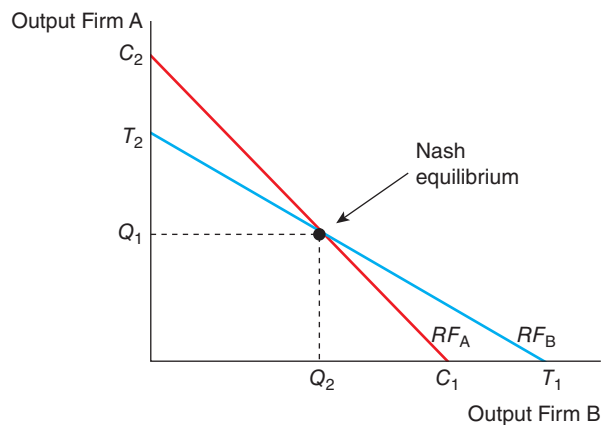
We can represent this equilibrium in Figure 16.10. Firm A's output is on the vertical axis and Firm B's output on the horizontal axis. If Firm A assumed Firm B would produce zero then it would supply the whole market and take all the profits given by point C_1 . If Firm B produced all the market output then Firm A would produce nothing indicated by point C_2 . All other points in-between show combinations of output for Firm A given a corresponding output decision by Firm B. The red line is Firm A's reaction function. The further away the output level from the whole market output the lower the firm's profits (and vice versa).

We can graph Firm B's reaction function in the same way and it will be the symmetrical opposite of Firm A's. If Firm A produces all the market output Firm B will produce nothing (point T_2) and if Firm A produced nothing Firm B would produce all the market output (point T_1). Connecting these points gives Firm B's reaction function indicated by the blue line.

FIGURE 16.10

Reaction Functions

The reaction functions given by the red and blue lines show combinations of output produced by Firm A and Firm B respectively given simultaneous decisions on output.



If Firm A assumed Firm B will produce all the output in the market it produces zero shown by the horizontal intercept C_1 . If Firm B produced nothing, Firm A will act as a monopolist and produce all the output shown by the vertical intercept C_2 . The points in-between show the various combinations of output Firm A would produce given Firm B's reaction and give Firm A's reaction function RF_A .

If Firm A produced nothing, Firm B would act as the monopolist and produce all the output shown by the horizontal intercept T_1 . If Firm A produced all the output Firm B would produce nothing shown by the vertical intercept T_2 . The points in-between give Firm B's reaction function shown by RF_B . If either firm produced off its reaction function at different time periods there would be an incentive for the other to change its output in subsequent time periods until equilibrium was reached at Q_1, Q_2 . This represents a Nash equilibrium because both firms make optimum decisions based on what its rivals are doing and a change in strategy by either firm would result in a less than optimum outcome.

The Bertrand Model

An alternative model is the Bertrand model developed by Joseph Bertrand in 1883. In this model firms' strategies are based on setting price – they are competing on price rather than on output. Again we assume a duopoly and that each firm has the same costs and produces a product which is highly substitutable – in other words the products are identical. Each firm assumes that the price fixed by its rivals is given. If Firm B, for example, sets a price, Firm A will select a price which maximizes profit and gives the output level to produce which maximizes this profit.

Assume that Firm B sets a price which is above MC . Firm A has the incentive to set its price slightly below that of Firm B and capture market share. Given that the products are identical it could capture all the market. In such a situation how would Firm B react? It would react by cutting its price. In turn Firm A would cut its price and the process will continue until $P = MC$ and there is no incentive for either to change. In such a situation it is rational for a firm to set price equal to marginal cost at the outset given what it expects its rival to do and of course the rival firm will think the same way. So, setting $P = MC$ is a Nash equilibrium.

What the Bertrand model predicts, therefore, is that the equilibrium in an oligopoly will give an outcome the same as that in perfect competition with $P = MC$. In such a situation both firms will be making its own pricing decision based on its own optimal outcome given the optimal behaviour of its rivals.

The Stackelberg Model

In reality, of course, firms do not make simultaneous decisions and game theory tells us that a firm will look to make decisions based on what it thinks its rivals will do in response and that it will learn from repeated instances of this 'game'. The Stackelberg model takes these new assumptions into account. This model looks at what equilibrium output would be for two firms if one firm was the leader taking decisions first but considering what the other firm would do in response. If Firm A announces that it is going to produce output Q_1 , it has to think about what Firm B will do in response. Firm A will have to consider a range of outputs and the response of Firm B to determine what its optimal output and profit level will be.

The Stackelberg model describes the advantages a firm can gain through moving before that of its rivals – so called first mover advantage. If Firm A is the leader and Firm B the follower then there may be an advantage to A in making a move to change output having estimated the response by Firm B. What is crucial in this model is that the decision made by Firm A is seen as being serious and credible by Firm B.

Assume that Firm A sets its output level at Q_1 , and thus is the first mover, Firm B, the follower, will observe this and make its output decision based on Firm A's decision. Firm A knows this and so will have to take into consideration what Firm B's reaction is going to be. It is in Firm A's interest to set output at a higher level to generate more profits but Firm B must also be sure that this is what Firm A is going to do hence the importance of Firm A's announcement being serious and credible. If Firm A announces that it intends to build a new plant and production facility to enable it to produce 70 per cent of the market output then Firm B must set its output to produce 30 per cent. If Firm B knows that Firm A is serious (and the investment announcement is an indicator of this) then if it produced any more than 30 per cent of the market output, price would be driven down due to excess supply and it would be worse off as a result.

In addition, moving first in the market could mean that Firm A is able to exploit the profits that accrue from its investment and use these profits to reinvest in expanding the firm further in the future. The risk for Firm A is that in making the move first it might experience considerable costs in so doing – costs which Firm B will also have to pay but possibly at lower levels because it can learn from the mistakes of the first mover.

PUBLIC POLICY TOWARD OLIGOPOLIES

One of the *Ten Principles of Economics* is that governments can sometimes improve market outcomes. The application of this principle to oligopolistic markets is, as a general matter, straightforward. As we have seen, cooperation among oligopolists is undesirable from the standpoint of society as a whole, because it leads to

production that is too low and prices that are too high. To move the allocation of resources closer to the social optimum, policymakers should try to induce firms in an oligopoly to compete rather than cooperate. Let's consider how policymakers do this and then examine the controversies that arise in this area of public policy.

Restraint of Trade and Competition Law

One way that policy discourages cooperation is through the common law. Normally, freedom of contract is an essential part of a market economy. Businesses and households use contracts to arrange mutually advantageous trades. In doing this, they rely on the court system to enforce contracts. Yet, for many centuries, courts in Europe and North America have deemed agreements among competitors to reduce quantities and raise prices to be contrary to the public interest. They have therefore refused to enforce such agreements.

Given the long experience of many European countries in tackling abuses of market power, it is perhaps not surprising that competition law is one of the few areas in which the European Union has been able to agree on a common policy. The European Commission can refer directly to the Treaty of Rome to prohibit price-fixing and other restrictive practices such as production limitation, and is especially likely to do so where a restrictive practice affects trade between EU member countries. The EU Competition Commission sets out its role as follows:

The antitrust area covers two prohibition rules set out in the Treaty on the Functioning of the European Union.

- *First, agreements between two or more firms which restrict competition are prohibited by Article 101 of the Treaty, subject to some limited exceptions. This provision covers a wide variety of behaviours. The most obvious example of illegal conduct infringing [the Article] is a cartel between competitors (which may involve price-fixing or market sharing).*
- *Second, firms in a dominant position may not abuse that position (Article 102 of the Treaty). This is, for example, the case for predatory pricing aiming at eliminating competitors from the market.*

The Commission is empowered by the Treaty to apply these prohibition rules and enjoys a number of investigative powers to that end (e.g. inspection in business and non-business premises, written requests for information, etc.). It may also impose fines on undertakings that violate EU antitrust rules. Since 1 May 2004, all national competition authorities are also empowered to apply fully the provisions of the Treaty in order to ensure that competition is not distorted or restricted. National courts may also apply these prohibitions so as to protect the individual rights conferred to citizens by the Treaty.

Controversies over Competition Policy

Over time, much controversy has centred on the question of what kinds of behaviour competition law should prohibit. Most commentators agree that price-fixing agreements among competing firms should be illegal. Yet competition law has been used to condemn some business practices whose effects are not obvious. Here we consider three examples.

Resale Price Maintenance One example of a controversial business practice is *resale price maintenance*, also called *fair trade*. Imagine that Superduper Electronics sells Blu-ray players to retail stores for €50. If Superduper requires the retailers to charge customers €75, it is said to engage in resale price maintenance. Any retailer that charged less than €75 would have violated its contract with Superduper.

At first, resale price maintenance might seem anti-competitive and, therefore, detrimental to society. Like an agreement among members of a cartel, it prevents the retailers from competing on price. For this reason, the courts have often viewed resale price maintenance as a violation of competition law.

Yet some economists defend resale price maintenance on two grounds. First, they deny that it is aimed at reducing competition. To the extent that Superduper Electronics has any market power, it can exert that power through the wholesale price, rather than through resale price maintenance. Moreover, Superduper has no incentive to discourage competition among its retailers. Indeed, because a cartel of retailers sells less than a group of competitive retailers, Superduper would be worse off if its retailers were a cartel.

Second, economists believe that resale price maintenance has a legitimate goal. Superduper may want its retailers to provide customers with a pleasant showroom and a knowledgeable sales force. Yet, without

resale price maintenance, some customers would take advantage of one store's service to learn about the Blu-ray player's special features and then buy the item at a discount retailer that does not provide this service. To some extent, good service is a public good among the retailers that sell Superduper products. When one person provides a public good, others are able to enjoy it without paying for it. In this case, discount retailers would free ride on the service provided by other retailers, leading to less service than is desirable. Resale price maintenance is one way for Superduper to solve this free-rider problem.

The example of resale price maintenance illustrates an important principle: business practices that appear to reduce competition may in fact have legitimate purposes. This principle makes the application of competition law all the more difficult. The competition authorities in each EU nation under the European Competition Network are in charge of enforcing these laws and must determine what kinds of behaviour public policy should prohibit as impeding competition and reducing economic well-being. Often that job is not easy.

Predatory Pricing Firms with market power normally use that power to raise prices above the competitive level. But should policymakers ever be concerned that firms with market power might charge prices that are too low? This question is at the heart of a second debate over competition policy.

Imagine that a large airline, call it Eurovia Airlines, has a monopoly on some route. Then Euro Express enters and takes 20 per cent of the market, leaving Eurovia with 80 per cent. In response to this competition, Eurovia starts slashing its fares. Some anti-trust analysts argue that Eurovia's move could be anti-competitive: the price cuts may be intended to drive Euro Express out of the market so Eurovia can recapture its monopoly and raise prices again. Such behaviour is called *predatory pricing*.

Although it is common for companies to complain to the relevant authorities that a competitor is pursuing predatory pricing, some economists are sceptical of this argument and believe that predatory pricing is rarely, and perhaps never, a profitable business strategy. Why? For a price war to drive out a rival, prices have to be driven below cost. Yet if Eurovia starts selling cheap tickets at a loss, it had better be ready to fly more planes, because low fares will attract more customers. Euro Express, meanwhile, can respond to Eurovia's predatory move by cutting back on flights. As a result, Eurovia ends up bearing more than 80 per cent of the losses, putting Euro Express in a good position to survive the price war.

Economists continue to debate whether predatory pricing should be a concern for competition policymakers. Various questions remain unresolved. Is predatory pricing ever a profitable business strategy? If so, when? Are the authorities capable of telling which price cuts are competitive and thus good for consumers and which are predatory? There are no simple answers.

Tying A third example of a controversial business practice is *tying*. Suppose that Makemoney Movies produces two new films – *Spiderman* and *Hamlet*. If Makemoney offers cinemas the two films together at a single price, rather than separately, the studio is said to be tying its two products.

Some economists have argued that the practice of tying should be banned. Their reasoning is as follows: imagine that *Spiderman* is a blockbuster, whereas *Hamlet* is an unprofitable art film. Then the studio could use the high demand for *Spiderman* to force cinemas to buy *Hamlet*. It seems that the studio could use tying as a mechanism for expanding its market power.

Other economists are sceptical of this argument. Imagine that cinemas are willing to pay €20,000 for *Spiderman* and nothing for *Hamlet*. Then the most that a cinema would pay for the two films together is €20,000 – the same as it would pay for *Spiderman* by itself. Forcing the cinema to accept a worthless film as part of the deal does not increase the cinema's willingness to pay. Makemoney cannot increase its market power simply by bundling the two films together.

Why, then, does tying exist? One possibility is that it is a form of price discrimination. Suppose there are two cinemas. City Cinema is willing to pay €15,000 for *Spiderman* and €5,000 for *Hamlet*. Country Cinema is just the opposite: it is willing to pay €5,000 for *Spiderman* and €15,000 for *Hamlet*. If Makemoney charges separate prices for the two films, its best strategy is to charge €15,000 for each film, and each cinema chooses to show only one film. Yet if Makemoney offers the two films as a bundle, it can charge each cinema €20,000 for the films. Thus, if different cinemas value the films differently, tying may allow the studio to increase profit by charging a combined price closer to the buyers' total willingness to pay.

Tying remains a controversial business practice. Microsoft has been investigated for 'tying' its internet browser and other software like its Windows Media Player with its Windows operating system.

The argument that tying allows a firm to extend its market power to other goods is not well founded, at least in its simplest form. Yet economists have proposed more elaborate theories for how tying can impede competition. Given our current economic knowledge, it is unclear whether tying has adverse effects for society as a whole.

All the analysis is based on an assumption that rivals may have sufficient information to be able to make a decision and that the decision will be a rational one based on this information. In reality firms do not have perfect information and do not behave rationally. Most firms in oligopolistic markets work very hard to protect sensitive information and only give out what they have to by law. Some information may be given to deliberately obfuscate the situation and hide what their true motives/strategies/tactics are. Economists have tried to include these imperfections into theories. Behavioural economics has become more popular in recent years because of the fact that it offers some greater insights into the **observed** behaviour of the real world which often does not conform to the assumptions of rationality.

SELF TEST What kind of agreement is illegal for businesses to make? • Why is competition law controversial?

CONCLUSION

Oligopolies would like to act like monopolies, but self-interest drives them closer to competition. Thus, oligopolies can end up looking either more like monopolies or more like competitive markets, depending on the number of firms in the oligopoly and how cooperative the firms are. The story of the prisoners' dilemma shows why oligopolies can fail to maintain cooperation, even when cooperation is in their best interest.

Policymakers regulate the behaviour of oligopolists through competition law. The proper scope of these laws is the subject of on-going controversy. Although price fixing among competing firms clearly reduces economic welfare and is declared as being illegal in many countries, some business practices that appear to reduce competition may have legitimate if subtle purposes. As a result, policymakers need to be careful when they use the substantial powers of competition law to place limits on firm behaviour.

IN THE NEWS



Collusion

In this chapter we have looked at oligopolies in the context of a relatively small number of firms in an industry which dominate the market. In this In the News, we look at how photocopying shops in one part of India are acting like an oligopoly by forming an association and colluding on price.

Xerox Stores in Bangalore

Oligopolistic behaviour can manifest itself in different ways. In parts of India, students rely heavily on photocopying materials for study. There

are two aspects to this practice. On the one hand, students claim that photocopying is the only way that they can access the study materials they need because they cannot afford to buy textbooks provided by

the major publishers. For publishers, the problem of students copying textbooks means that the sales of textbooks are falling. Not only does this affect sales but it also affects intellectual property rights. Somewhat

perversely, photocopying textbooks increases publishers' costs as they divert resources to fight the problem and this means that prices are forced upwards. The cycle continues. Attempts to reduce the price of textbooks in parts of India does not seem to have helped matters. One bookshop owner is reported to have noted that some textbooks priced at Rs 5,000 (about €70) were reduced in price to nearer Rs 500 (about €7) but even this price reduction was not enough to encourage a significant increase in sales such is the habitual nature of photocopying.

Businesses set up to provide photocopying services, known as Xerox stores, typically provide the means by which copies of textbooks can be made available to students at under Rs 200 (around €2.75). The situation, however, might change in one area, Bangalore, where photocopying prices are on the rise. Photocopying businesses in the area have formed the Bangalore Owners' Welfare Association (BOWA) and have been attempting to show customers problems they are facing with rising costs.

Over the last 20 years, the price of paper has risen by 275 per cent, toner is up by 2,300 per cent per kilo, electricity has risen in price seven times over the period and fuel has risen by 252 per cent. These increases in cost have not been matched, claim the owners, by equivalent rises in the prices people are paying for photocopies; indeed, some the owners point to the fact that the price has fallen since 1988 by 66 per cent and at the end of 2012 stood at around 35 paise (there are 100 paise to a rupee) – around €0.005.

Individual shop owners have wanted to increase prices for a number of years but felt constrained because of how other owners might respond. If one owner put prices up they risked losing lots of business to other Xerox stores. The result of this situation was that no one put prices up. Cost pressures have meant that owners decided to get together to form the BOWA and one of the first actions the Association has made is to announce a price increase. Individual copies will rise to Rs 2 (€0.03) and if customers want to

photocopy bulk orders (like textbooks) the price is going up by up to 186 per cent. The BOWA has around 2,800 members and it has been made mandatory that every shop increases its prices.

The rise in the price of photocopying has not been welcomed by students and their representatives. One of the reasons for the low price of copies at shops and the reluctance of owners to increase price was because students had been arriving at stores with large orders and demanding price discounts for bulk photocopying. The rise in price will now mean that photocopying textbooks is not quite as cost-effective as it once was, but student groups are claiming that they are being unfairly treated because they do not have an alternative means of accessing the study materials they need.

Questions

- 1 The textbook publishing industry is dominated by a relatively small number of large firms. To what extent is the publishing industry an oligopoly in India? Explain your reasoning.
- 2 In what way is photocopying of textbooks an example of entry to the market and how might publishers go about erecting more effective barriers to entry?
- 3 How did the behaviour of shop owners prior to the formation of the BOWA exhibit the characteristic of interdependence?
- 4 To what extent would you argue that the BOWA is an example of a cartel and as a result should be declared anti-competitive?
- 5 How long do you think the BOWA can maintain the price increase it has ordered shops to make? Explain with reference to the content in this chapter.

Photocopying shops in one part of India are acting like an oligopoly by forming an association and colluding on price.



SUMMARY

- Oligopolists maximize their total profits by forming a cartel and acting like a monopolist. Yet, if oligopolists make decisions about production levels individually, the result is a greater quantity and a lower price than under the monopoly outcome. The larger the number of firms in the oligopoly, the closer the quantity and price will be to the levels that would prevail under competition.
- The prisoners' dilemma shows that self-interest can prevent people from maintaining cooperation, even when cooperation is in their mutual interest. The logic of the prisoners' dilemma applies in many situations, including advertising, common-resource problems and oligopolies.
- There are different models of oligopoly which each make different assumptions and result in different market outcomes.
- Policymakers use competition law to prevent oligopolies from engaging in behaviour that reduces competition. The application of these laws can be controversial, because some behaviour that may seem to reduce competition may in fact have legitimate business purposes.

QUESTIONS FOR REVIEW

- 1 What is meant by the term concentration ratio? If three firms account for 72 per cent of total market sales what is the concentration ratio?
- 2 What are the main characteristics of an oligopolistic market structure?
- 3 If a group of sellers could form a cartel, what quantity and price would they try to set?
- 4 Compare the quantity and price of an oligopoly with those of a monopoly and a competitive market.
- 5 How does the number of firms in an oligopoly affect the outcome in its market?
- 6 What is the kinked demand curve and why might it be a feature of an oligopolistic market?
- 7 What is the prisoners' dilemma, and what does it have to do with oligopoly?
- 8 What effect might repeated playing of the prisoners' dilemma have on the equilibrium of an oligopoly?
- 9 Explain the difference in approach of the Cournot and Bertrand models of oligopoly.
- 10 What kinds of behaviour do the competition laws prohibit?

PROBLEMS AND APPLICATIONS

- 1 A large share of the world supply of diamonds comes from Russia and South Africa. Suppose that the marginal cost of mining diamonds is constant at €1,000 per diamond, and the demand for diamonds is described by the following schedule:

| Price (€) | Quantity |
|-----------|----------|
| 8,000 | 5,000 |
| 7,000 | 6,000 |
| 6,000 | 7,000 |
| 5,000 | 8,000 |
| 4,000 | 9,000 |
| 3,000 | 10,000 |
| 2,000 | 11,000 |
| 1,000 | 12,000 |

- a. If there were many suppliers of diamonds, what would be the price and quantity?
 - b. If there was only one supplier of diamonds, what would be the price and quantity?
 - c. If Russia and South Africa formed a cartel, what would be the price and quantity? If the countries split the market evenly, what would be South Africa's production and profit? What would happen to South Africa's profit if it increased its production by 1,000 while Russia stuck to the cartel agreement?
 - d. Use your answer to part (c) to explain why cartel agreements are often not successful.
- 2 This chapter discusses companies that are oligopolists in the market for the goods they sell. Many of the same ideas apply to companies that are oligopolists in the market for the inputs they buy. If sellers who are oligopolists try to increase the price of goods they sell, what is the goal of buyers who are oligopolists?
 - 3 Describe several activities in your life in which game theory could be useful. What is the common link among these activities?

4 Suppose that you and a fellow student are assigned a project on which you will receive one combined grade. You each want to receive a good grade (which means you have to work), but you also want to do as little work as possible (which means you shirk).

In particular, here is the situation:

- If both of you work hard, you both get an A, which gives each of you 40 units of happiness.
- If only one of you works hard, you both get a B, which gives each of you 30 units of happiness.
- If neither of you works hard, you both get a D, which gives each of you 10 units of happiness.

a. Fill in the payoffs in the following matrix:

| | | | |
|----------------------|-------|--------------------|--------------------|
| | | Your decision | |
| | | Work | Shirk |
| Classmate's decision | Work | You: Classmate: | You: Classmate: |
| | Shirk | You: Classmate: | You: Classmate: |

- b. What is the likely outcome? Explain your answer.
- c. If you get this person as your partner on a series of projects throughout the year, rather than only once, how might that change the outcome you predicted in part b?
- d. Another person on your course cares more about good grades. He gets 50 units of happiness for a B and 80 units of happiness for an A. If this person was your partner (but your preferences were unchanged), how would your answers to parts a. and b. change? Which of the two partners would you prefer? Would he also want you as a partner?

5 Synergy and Dynaco are the only two firms in a specific hi-tech industry. They face the following payoff matrix as they decide upon the size of their research budget.

| | | | |
|-------------------|--------------|---|---|
| | | Synergy's decision | |
| | | Large budget | Small budget |
| Dynaco's decision | Large budget | Synergy gains \$20 million Dynaco gains \$30 million | Synergy gains zero Dynaco gains \$70 million |
| | Small budget | Synergy gains \$30 million Dynaco gains zero | Synergy gains \$40 million Dynaco gains \$50 million |

- a. Does Synergy have a dominant strategy? Explain.
- b. Does Dynaco have a dominant strategy? Explain.
- c. Is there a Nash equilibrium for this scenario? Explain.

6 In the 1970s, concern over the health effects of tobacco led to many countries banning tobacco advertising on television.

- a. Why might tobacco companies have not fought too hard against the ban?
- b. In the wake of the ban, the profits of tobacco companies rose. Why might this scenario have occurred?
- c. Could the ban still be good public policy even though tobacco company profits grew? Explain your answer.

7 Assume that two airline companies decide to engage in collusive behaviour.

Let's analyse the game between two such companies. Suppose that each company can charge either a high price for tickets or a low price. If one company charges €100, it earns low profits if the other company charges €100 also, and high profits if the other company charges €200. On the other hand, if the company charges €200, it earns very low profits if the other company charges €100, and medium profits if the other company charges €200 also.

- a. Draw the payoff matrix for this game.
- b. What is the Nash equilibrium in this game? Explain.
- c. Is there an outcome that would be better than the Nash equilibrium for both airlines? How could it be achieved? Who would lose if it were achieved?

8 Farmer Jones and Farmer MacDonald graze their cattle on the same field. If there are 20 cows grazing in the field, each cow produces €4,000 of milk over its lifetime. If there are more cows in the field, then each cow can eat less grass, and its milk production falls. With 30 cows on the field, each produces €3,000 of milk; with 40 cows, each produces €2,000 of milk. Cows cost €1,000 apiece.

- a. Assume that Farmer Jones and Farmer MacDonald can each purchase either ten or 20 cows, but that neither knows how many the other is buying when she makes her purchase. Calculate the payoffs of each outcome.
- b. What is the likely outcome of this game? What would be the best outcome? Explain.
- c. There used to be more common fields than there are today. Why?

9 Little Kona is a small coffee company that is considering entering a market dominated by Big Brew. Each company's profit depends on whether

Little Kona enters and whether Big Brew sets a high price or a low price:

| | | | |
|-------------|-------------|--|--|
| | | Big Brew | |
| | | High price | Low price |
| Little Kona | Enter | Brew makes €3 million Kona makes €2 million | Brew makes €1 million Kona loses €1 million |
| | Don't enter | Brew makes €7 million Kona makes zero | Brew makes €2 million Kona makes zero |

Big Brew threatens Little Kona by saying, 'If you enter, we're going to set a low price, so you had better stay out.' Do you think Little Kona should believe the threat? Why or why not? What do you think Little Kona should do?

- 10** Consider an oligopoly with four firms who decide to engage in a price war. Can there be any winners in this war? What is the relevance of the Bertrand model of oligopoly and the notion of repeated games to the strategy of a price war?

PART 7

FACTOR MARKETS

17

THE ECONOMICS OF LABOUR MARKETS

When you finish university, your income will be determined largely by what kind of job you take. If you become a professional economist, you will earn more than if you become a petrol station attendant. This fact is not surprising, but it is not obvious why it is true. No law requires that economists be paid more than petrol station attendants. No ethical principle says that economists are more deserving. If you become an academic economist in a university, you will probably earn less than if you become a banker. What then determines which job will pay you the higher wage?

People earn income in various ways. Wages, salaries and fringe benefits such as pension provision, health insurance and bonuses form the majority of income for labour in most economies but income can also be earned from self-employment.

The amount people earn from selling their labour, like many questions in economics, hinges on supply and demand. The supply and demand for labour determine the prices paid to workers. To understand why some people have higher incomes than others, therefore, we need to look more deeply at the markets for the services they provide.

This chapter provides the basic theory for the analysis of factor markets in general. In the next chapter we will look at the market for capital and land. Although in many ways factor markets resemble the goods markets we have analysed in previous chapters, they are different in one important way: the demand for a factor of production is a **derived demand**. That is, a firm's demand for a factor of production is derived (determined) from its decision to supply a good in another market. The demand for computer programmers is inextricably tied to the supply of computer software, and the demand for petrol station attendants is inextricably tied to the supply of petrol.

derived demand a situation where demand is determined by the supply in another market

The initial analysis will be based on firms operating in a competitive market – both for goods and labour. In the same way that we assumed free entry and exit of firms in perfect competition, we will base the analysis on the assumption that labour is free to enter and exit the market, and firms are equally free to employ and shed labour at will – in other words, people can move into and out of work easily and

employers can ‘hire and fire’ workers when they need to. In reality, of course, there are a number of imperfections in the labour market which we will analyse later in the chapter but our initial analysis serves to act as a benchmark for looking at how labour markets work in reality.

THE DEMAND FOR LABOUR

To understand labour demand, we need to focus on the firms that hire the labour and use it to produce goods for sale. As we have already noted, labour markets are different from most other markets because labour demand is a derived demand. Most labour services, rather than being final goods ready to be enjoyed by consumers, are inputs into the production of other goods. Firms hire workers because workers contribute to production and the payment to workers (the wage) is the price that employers have to pay to hire those labour services. Understanding that the demand for labour is inextricably linked to the ability of firms to sell the goods and services which labour helps produce is fundamental to understanding labour markets. By examining the link between the production of goods and the demand for labour, we gain insight into the determination of equilibrium wages.

The Competitive Profit-Maximizing Firm

Let’s look at how a typical firm, we will use the example of an apple producer to help illustrate the principles, decides the quantity of labour to demand. The firm owns an apple orchard and during the harvesting period must decide how many apple pickers to hire to pick the crop. After the firm makes its hiring decision, the workers pick as many apples as they can. The firm then sells the apples, pays the workers, and keeps what is left as profit.

Assumption 1: Competitive Conditions We assume that our firm is *competitive* both in the market for apples (where the firm is a seller) and in the market for apple pickers (where the firm is a buyer). Because there are many other firms selling apples and hiring apple pickers, a single firm has little influence over the price it gets for apples or the wage it pays apple pickers. The firm takes the price and the wage as given by market conditions. It only has to decide how many workers to hire and how many apples to sell.

Assumption 2: Firms Are Profit Maximizers Second, we assume that the firm is *profit-maximizing*. Thus, the firm does not directly care about the number of workers it has or the number of apples it produces. It cares only about profit, which equals the total revenue from the sale of apples minus the total cost of producing them. The firm’s supply of apples and its demand for workers are derived from its primary goal of maximizing profit.

The Production Function and the Marginal Product of Labour

To make its hiring decision, the firm must consider how the number of apple pickers affects the quantity of apples it can harvest and sell. Table 17.1 gives a numerical example. In the first column is the number of workers. In the second column is the quantity of apples the workers harvest each week.

These two columns of numbers describe the firm’s ability to produce. Remember that the production function describes the relationship between the quantity of the inputs used in production and the quantity of output from production. Here the ‘input’ is the apple pickers and the ‘output’ is the apples. The other inputs – the trees themselves, the land, the firm’s trucks and tractors, and so on – are held fixed for now. This firm’s production function shows that if the firm hires one worker, that worker will pick 1,000 kilos of apples per week. If the firm hires two workers, the two workers together will pick 1,800 kilos per week, and so on.

Figure 17.1 graphs the data on labour and output presented in Table 17.1. The number of workers is on the horizontal axis, and the amount of output is on the vertical axis. This figure illustrates the production function.

TABLE 17.1

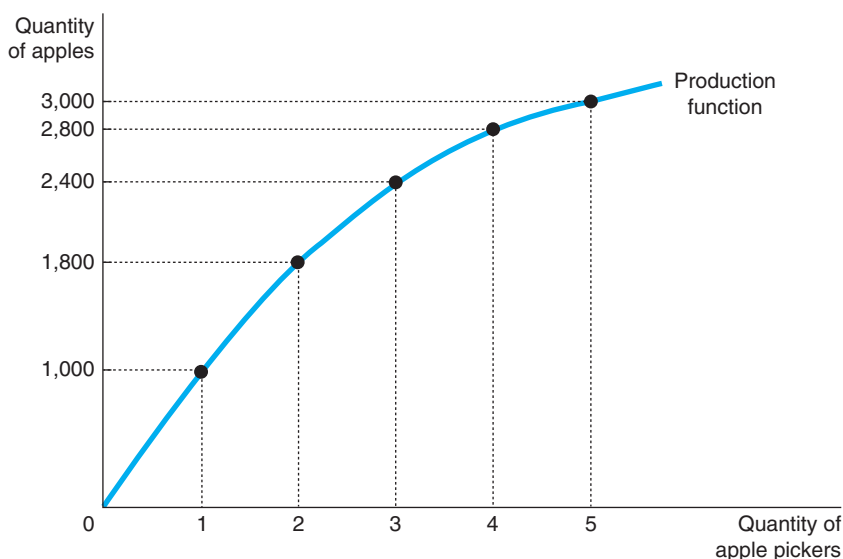
How the Competitive Firm Decides How Much Labour to Hire

| Labour | Output | Marginal product of labour | Value of the marginal product of labour | Wage | Marginal profit |
|----------------------------|-------------------------|--|---|---------|--|
| L (number of workers) | Q (kilos per week) | $MP_L = \Delta Q / \Delta L$ (kilos per week) | $VMP_L = P \times MP_L$ (€) | W (€) | $\Delta \text{Profit} = VMP_L - W$ (€) |
| 0 | 0 | | 1,000 | 500 | |
| 1 | 1,000 | 1,000 | 800 | 500 | 500 |
| 2 | 1,800 | 800 | 600 | 500 | 300 |
| 3 | 2,400 | 600 | 400 | 500 | 100 |
| 4 | 2,800 | 400 | 200 | 500 | -100 |
| 5 | 3,000 | 200 | | | -300 |

FIGURE 17.1

The Production Function

The production function is the relationship between the inputs into production (apple pickers) and the output from production (apples). As the quantity of the input increases, the production function gets flatter, reflecting the property of diminishing marginal product.



The principle that rational people think at the margin is key to understanding how firms decide what quantity of labour to hire. To take a step towards this decision, the third column in Table 17.1 gives the **marginal product of labour**, the increase in the amount of output from an additional unit of labour. When the firm increases the number of workers from 1 to 2, for example, the amount of apples produced rises from 1,000 to 1,800 kilos. Therefore, the marginal product of the second worker is 800 kilos.

marginal product of labour the increase in the amount of output from an additional unit of labour

Notice that as the number of workers increases and because other factors such as land and capital are fixed, the marginal product of labour declines because of the principle of diminishing marginal product. At first, when only a few workers are hired, they pick apples from the best trees in the orchard. As the number of workers increases, additional workers have to pick from the trees with fewer apples. Hence, as more and more workers are hired, each additional worker contributes less to the production of apples. For this reason, the production function in Figure 17.1 becomes flatter as the number of workers rises.

The Value of the Marginal Product and the Demand for Labour

Our profit-maximizing firm is concerned more with money than with apples. As a result, when deciding how many workers to hire, the firm considers how much profit each worker would bring in. Because profit is total revenue minus total cost, the profit from an additional worker is the worker's contribution to revenue minus the worker's wage.

To find the worker's contribution to revenue, we must convert the marginal product of labour (which is measured in kilos of apples) into the *value* of the marginal product (which is measured in euros). We do this using the price of apples. If a kilo of apples sells for €1 and if an additional worker produces 800 kilos of apples, then the worker produces €800 of revenue for the firm.

The **value of the marginal product** of any input is the marginal product of that input multiplied by the market price of the output. You might also see this referred to as the *marginal physical product* (MPP). The fourth column in Table 17.1 shows the value of the marginal product of labour in our example, assuming the price of apples is €1 per kilo. Because the market price is constant for a competitive firm, the value of the marginal product (like the marginal product itself) diminishes as the number of workers rises. Economists sometimes call this column of numbers the firm's **marginal revenue product**: it is the extra revenue the firm gets from hiring an additional unit of a factor of production.

value of the marginal product the marginal product of an input times the price of the output
marginal revenue product the extra revenue a firm gets from hiring an additional unit of a factor of production

Now consider how many workers the firm will hire. Suppose that the market wage for apple pickers is €500 per week. In this case, as you see in Table 17.1, the first worker that the firm hires is profitable: the first worker yields €1,000 in revenue, or €500 in profit. Similarly, the second worker yields €800 in additional revenue, or €300 in profit. The third worker produces €600 in additional revenue, or €100 in profit. After the third worker, however, hiring workers is unprofitable. The fourth worker would yield only €400 of additional revenue. Because the worker's wage is €500, hiring the fourth worker would mean a €100 reduction in profit. Thus, the firm hires only three workers.

It is instructive to consider the firm's decision graphically. Figure 17.2 graphs the value of the marginal product. This curve slopes downward because the marginal product of labour diminishes as the number of workers rises. The figure also includes a horizontal line at the market wage. To maximize profit, the firm hires workers up to the point where these two curves cross. Below this level of employment, the value of the marginal product exceeds the wage, so hiring another worker would increase profit. Above this level of employment, the value of the marginal product is less than the wage, so the marginal worker is unprofitable. Thus, a competitive, profit-maximizing firm hires workers up to the point where the value of the marginal product of labour equals the wage.

Having explained the profit-maximizing hiring strategy for a competitive firm, we can now offer a theory of labour demand. Recall that a firm's labour demand curve tells us the quantity of labour that a firm demands at any given wage. We have just seen in Figure 17.2 that the firm makes that decision by choosing the quantity of labour at which the value of the marginal product equals the wage. As a result, the value of marginal product curve is the labour demand curve for a competitive, profit-maximizing firm.

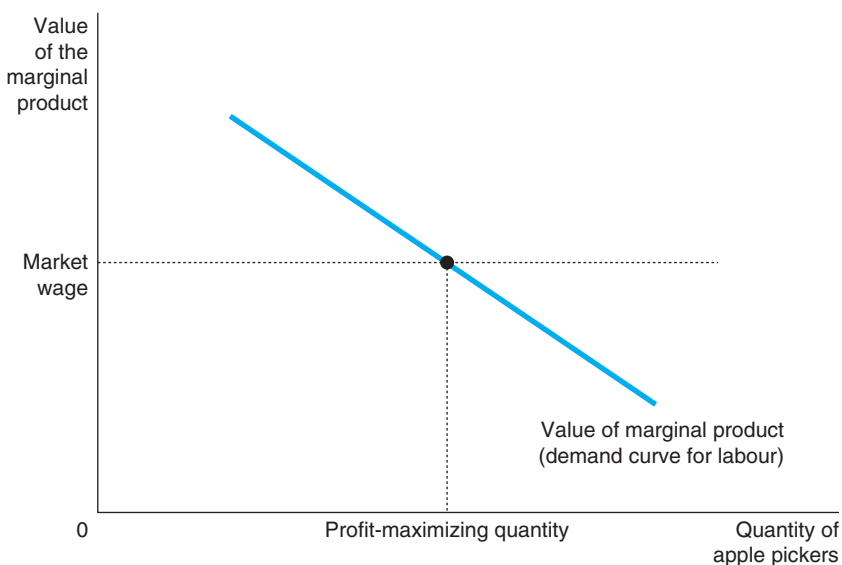
Input Demand and Output Supply: Two Sides of the Same Coin

We know that a competitive, profit-maximizing firm decides how much of its output to sell: it chooses the quantity of output at which the price of the good equals the marginal cost of production. We have just seen that a firm chooses the quantity of labour at which the wage equals the value of the marginal product. Because the production function links the quantity of inputs to the quantity of output, you should not be surprised to learn that the firm's decision about input demand is closely linked to its decision about output supply. In fact, these two decisions are two sides of the same coin.

To see this relationship more fully, let's consider how the marginal product of labour (MP_L) and marginal cost (MC) are related. Suppose an additional worker costs €500 and has a marginal product of 50 kilos of apples. In this case, producing 50 more kilos costs €500; the marginal cost of a kilo is €500/50, or €10.

FIGURE 17.2**The Value of the Marginal Product of Labour**

This figure shows how the value of the marginal product (the marginal product times the price of the output) depends on the number of workers. The curve slopes downward because of diminishing marginal product. For a competitive, profit-maximizing firm, this value of marginal product curve is also the firm's labour demand curve.



More generally, if W is the wage, and an extra unit of labour produces MP_L units of output, then the marginal cost of a unit of output is $MC = W/MP_L$.

This analysis shows that diminishing marginal product is closely related to increasing marginal cost. When our apple orchard grows crowded with workers, each additional worker adds less to the production of apples (MP_L falls). Similarly, when the apple firm is producing a large quantity of apples, the orchard is already crowded with workers, so it is more costly to produce an additional kilo of apples (MC rises).

Now consider our criterion for profit maximization. We determined earlier that a profit-maximizing firm chooses the quantity of labour so that the value of the marginal product ($P \times MP_L$) equals the wage (W). We can write this mathematically as:

$$P \times MP_L = W$$

If we divide both sides of this equation by MP_L , we obtain:

$$P = W/MP_L$$

We just noted that W/MP_L equals marginal cost MC . Therefore, we can substitute to obtain:

$$P = MC$$

This equation states that the price of the firm's output is equal to the marginal cost of producing a unit of output. *Thus, when a competitive firm hires labour up to the point at which the value of the marginal product equals the wage, it also produces up to the point at which the price equals marginal cost.* Our analysis of labour demand in this chapter is just another way of looking at the production decision we first encountered in Chapter 6.

What Causes the Labour Demand Curve to Shift?

The labour demand curve reflects the value of the marginal product of labour. With this insight in mind, let's consider a few of the things that might cause the labour demand curve to shift.

The Output Price The value of the marginal product is marginal product times the price of the firm's output. Thus, when the output price changes, the value of the marginal product changes, and the labour demand curve shifts. An increase in the price of apples, for instance, raises the value of the marginal

product of each worker who picks apples and, therefore, increases labour demand from the firms that supply apples. Conversely, a decrease in the price of apples reduces the value of the marginal product and decreases labour demand.

Technological Change Over the last 20 years labour productivity, the amount produced per hour per worker, in the EU has risen. The most important reason why is technological progress: scientists and engineers are constantly figuring out new and better ways of doing things. This has profound implications for the labour market. Technological advance raises the marginal product of labour, which in turn increases the demand for labour. Such technological advance explains persistently rising employment in the face of rising wages.

The Supply of Other Factors The quantity available of one factor of production can affect the marginal product of other factors. A fall in the supply of ladders, for instance, will reduce the marginal product of apple pickers and thus the demand for apple pickers.

SELF TEST Define *marginal product of labour* and *value of the marginal product of labour*. • Describe how a competitive, profit-maximizing firm decides how many workers to hire.

THE SUPPLY OF LABOUR

The other side of the market is labour supply. Individuals offer their labour services in return for payment (wages and salaries).

The Trade-Off between Work and Leisure

One of the *Ten Principles of Economics* is that people face trade-offs. Probably no trade-off is more obvious or more important in a person's life than the trade-off between work and leisure. The more hours you spend working, the fewer hours you have to watch TV, socialize with friends or pursue your favourite hobby. The trade-off between labour and leisure lies behind the labour supply curve.

Another of the *Ten Principles of Economics* is that the cost of something is what you give up to get it. What do you give up to get an hour of leisure? You give up an hour of work, which in turn means an hour of wages. Thus, if your wage is €15 per hour, the opportunity cost of an hour of leisure is €15. And if you get a pay rise to €20 per hour, the opportunity cost of enjoying leisure goes up.

The labour supply curve reflects how workers' decisions about the labour–leisure trade-off respond to a change in that opportunity cost. An upward sloping labour supply curve means that an increase in the wage induces workers to increase the quantity of labour they supply. Because time is limited, more hours of work mean that workers are enjoying less leisure. That is, workers respond to the increase in the opportunity cost of leisure by taking less of it.

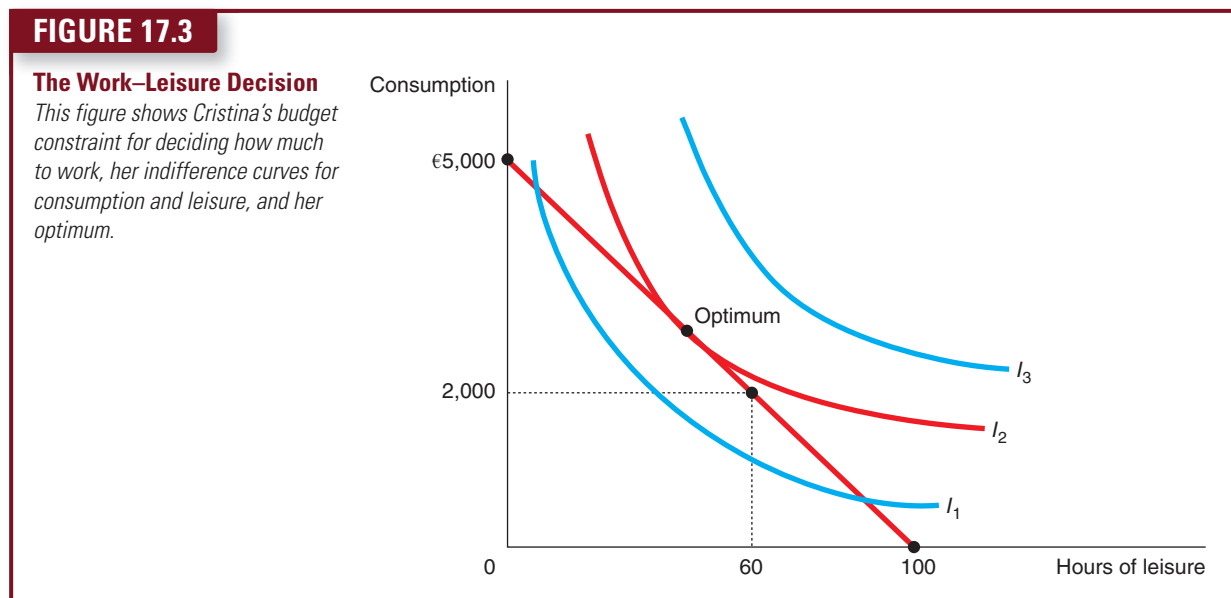
How Do Wages Affect Labour Supply?

In Chapter 5 we looked at the concept of income and substitution effects in analysing how a person decides how to allocate his income between two goods. We can use the same theory to analyse how a person decides to allocate his time between work and leisure.

Consider the decision facing Cristina, a freelance software designer. Cristina is awake for 100 hours per week. She spends some of this time enjoying leisure – going out with friends, watching television, going to the cinema and night clubs, and so on. She spends the rest of this time developing software at her computer. For every hour she spends developing software, she earns €50, which she spends on consumption goods. Thus, her wage (€50) reflects the trade-off Cristina faces between leisure and consumption. For every hour of leisure she gives up, she works one more hour and gets €50 of consumption.

Figure 17.3 shows Cristina's budget constraint. If she spends all 100 hours enjoying leisure, she has no consumption. If she spends all 100 hours working, she earns a weekly consumption of €5,000 but has no time for leisure. If she works a normal 40-hour week, she enjoys 60 hours of leisure and has weekly consumption of €2,000.

Figure 17.3 uses indifference curves to represent Cristina's preferences for consumption and leisure. Here consumption and leisure are the two 'goods' between which Cristina is choosing. Because Cristina always prefers more leisure and more consumption, she prefers points on higher indifference curves to points on lower ones. At a wage of €50 per hour, Cristina chooses a combination of consumption and leisure represented by the point labelled 'optimum'. This is the point on the budget constraint that is on the highest possible indifference curve, which is curve I_2 .



Now consider what happens when Cristina's wage increases from €50 to €60 per hour. Figure 17.4 shows two possible outcomes. In each case, the budget constraint, shown in the left-hand graph, pivots outward from BC_1 to BC_2 . In the process, the budget constraint becomes steeper, reflecting the change in relative price: at the higher wage, Cristina gets more consumption for every hour of leisure that she gives up.

Cristina's preferences, as represented by her indifference curves, determine the resulting responses of consumption and leisure to the higher wage. In both panels, consumption rises. Yet the response of leisure to the change in the wage is different in the two cases. In panel (a), Cristina responds to the higher wage by enjoying less leisure. In panel (b), Cristina responds by enjoying more leisure.

Cristina's decision between leisure and consumption determines her supply of labour because the more leisure she enjoys, the less time she has left to work. In each panel, the right-hand graph in Figure 17.4 shows the labour supply curve implied by Cristina's decision. In panel (a), a higher wage induces Cristina to enjoy less leisure and work more, so the labour supply curve slopes upward. In panel (b), a higher wage induces Cristina to enjoy more leisure and work less, so the labour supply curve slopes 'backwards'.

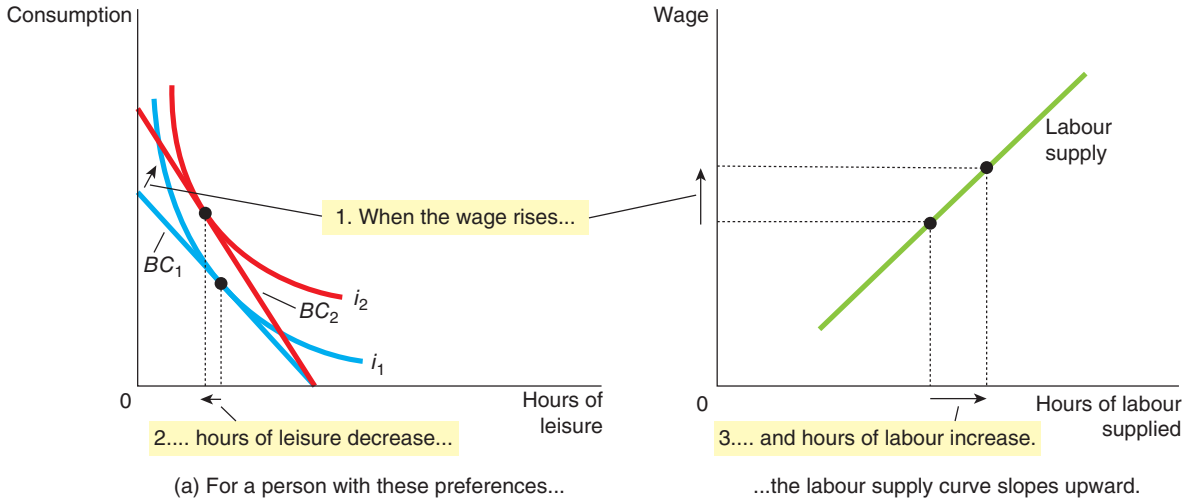
At first, the backwards sloping labour supply curve is puzzling. Why would a person respond to a higher wage by working less? The answer comes from considering the income and substitution effects of a higher wage.

Consider first the substitution effect. When Cristina's wage rises, leisure becomes more costly relative to consumption, and this encourages Cristina to substitute consumption for leisure. In other words, the substitution effect induces Cristina to work harder in response to higher wages, which tends to make the labour supply curve slope upward.

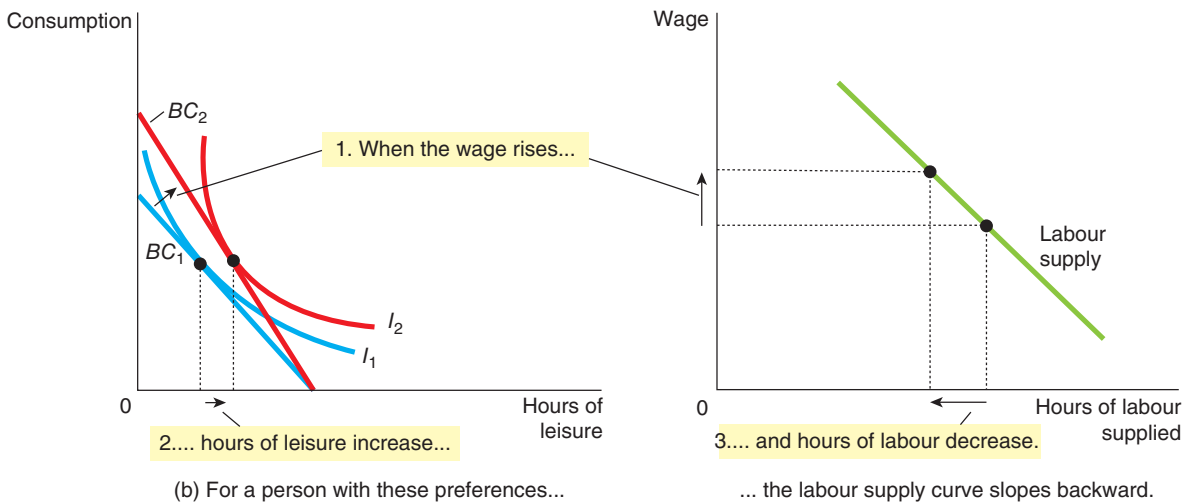
Now consider the income effect. When Cristina's wage rises, she moves to a higher indifference curve. She is now better off than she was. As long as consumption and leisure are both normal goods, she tends to want to use this increase in well-being to enjoy both higher consumption and greater leisure. In other words, the income effect induces her to work less, which tends to make the labour supply curve slope backwards.

FIGURE 17.4**An Increase in the Wage**

The two panels of this figure show how a person might respond to an increase in the wage. The graphs on the left show the consumer's initial budget constraint BC_1 , and new budget constraint BC_2 , as well as the consumer's optimal choices over consumption and leisure. The graphs on the right show the resulting labour supply curve. Because hours worked equal total hours available minus hours of leisure, any change in leisure implies an opposite change in the quantity of labour supplied. In panel (a), when the wage rises, consumption rises and leisure falls, resulting in a labour supply curve that slopes upwards. In panel (b), when the wage rises, both consumption and leisure rise, resulting in a labour supply curve that slopes backwards.



(a) For a person with these preferences...



(b) For a person with these preferences...

In the end, economic theory does not give a clear prediction about whether an increase in the wage induces Cristina to work more or less. If the substitution effect is greater than the income effect for Cristina, she works more. If the income effect is greater than the substitution effect, she works less. The labour supply curve, therefore, could be either upwards or backwards sloping. This concept has an important application to debates over the effect of tax cuts on work. Some economists argue that cutting taxes encourages people to work more hours because the reward is greater. Such an argument is also used as the basis for supporting an entrepreneurial culture – keep taxes low and this encourages entrepreneurs. Others point out that lower taxes do increase disposable income but workers may now use this higher income to enjoy more leisure and not work additional hours. Having some idea of the relative strength of the income and substitution effects is important in analysing and assessing such policy initiatives.

CASE STUDY

Income Effects on Labour Supply: Historical Trends, Lottery Winners and the Carnegie Conjecture

The idea of a backwards sloping labour supply curve might at first seem like a mere theoretical curiosity, but in fact it is not. Evidence indicates that the labour supply curve, considered over long periods of time, does in fact slope backwards. A hundred years ago many people in Europe and North America worked six days a week. Today five-day working weeks are the norm. At the same time that the length of the working week has been falling, the wage of the typical worker (adjusted for inflation) has been rising.

Here is how economists explain this historical pattern: over time, advances in technology raise workers' productivity and, thereby, the demand for labour. The increase in labour demand raises equilibrium wages. As wages rise, so does the reward for working. Yet rather than responding to this increased incentive by working more, most workers choose to take part of their greater prosperity in the form of more leisure. In other words, the income effect of higher wages dominates the substitution effect.

Further evidence that the income effect on labour supply is strong comes from a very different kind of data: winners of lotteries. Winners of large prizes in the lottery see large increases in their incomes and, as a result, large outward shifts in their budget constraints. Because the winners' wages have not changed, however, the *slopes* of their budget constraints remain the same. There is, therefore, no substitution effect. By examining the behaviour of lottery winners, we can isolate the income effect on labour supply. Nearly all the research on the effects of winning the lottery on labour supply has so far been done in the USA, but the results are striking. Of those winners who win more than \$50,000, almost 25 per cent quit working within a year, and another 9 per cent reduce the number of hours they work. Of those winners who win more than \$1 million, almost 40 per cent stop working. The income effect on labour supply of winning such a large prize is substantial.

Similar results were found in a study, published in the May 1993 issue of the *Quarterly Journal of Economics*, of how receiving a bequest affects a person's labour supply. The study found that a single person who inherits more than \$150,000 is four times as likely to stop working as a single person who inherits less than \$25,000. This finding would not have surprised the 19th-century industrialist Andrew Carnegie. Carnegie warned that 'the parent who leaves his son enormous wealth generally deadens the talents and energies of the son, and tempts him to lead a less useful and less worthy life than he otherwise would'. That is, Carnegie viewed the income effect on labour supply to be substantial and, from his paternalistic perspective, regrettable. During his life and at his death, Carnegie gave much of his vast fortune to charity.

If you won the lottery would you substitute more leisure for work?



What Causes the Labour Supply Curve to Shift?

The labour supply curve shifts whenever people change the amount they want to work at a given wage. Let's now consider some of the events that might cause such a shift.

Changes in Tastes A generation or two ago, it was the norm for women to stay at home while raising children. Today, family sizes are smaller and more mothers choose to work. The result is an increase in the supply of labour.

Changes in Alternative Opportunities The supply of labour in any one labour market depends on the opportunities available in other labour markets. If the wage earned by pear pickers suddenly rises, some apple pickers may choose to switch occupations. The supply of labour in the market for apple pickers falls.

Immigration Movement of workers from region to region, or country to country, is an obvious and often important source of shifts in labour supply. When immigrants move from one European country to another – from Poland to the UK, for instance – the supply of labour in the United Kingdom increases and the supply of labour in Poland contracts. In fact, much of the policy debate about immigration centres on its effect on labour supply and, thereby, equilibrium in the labour market.

SELF TEST Who has a greater opportunity cost of enjoying leisure – a petrol station attendant or a brain surgeon? Explain. Can this help explain why doctors work such long hours?

EQUILIBRIUM IN THE LABOUR MARKET

We can establish two facts about how wages are determined in competitive labour markets:

- The wage adjusts to balance the supply and demand for labour.
- The wage equals the value of the marginal product of labour.

At first, it might seem surprising that the wage can do both these things at once. In fact, there is no real puzzle here, but understanding why there is no puzzle is an important step to understanding wage determination.

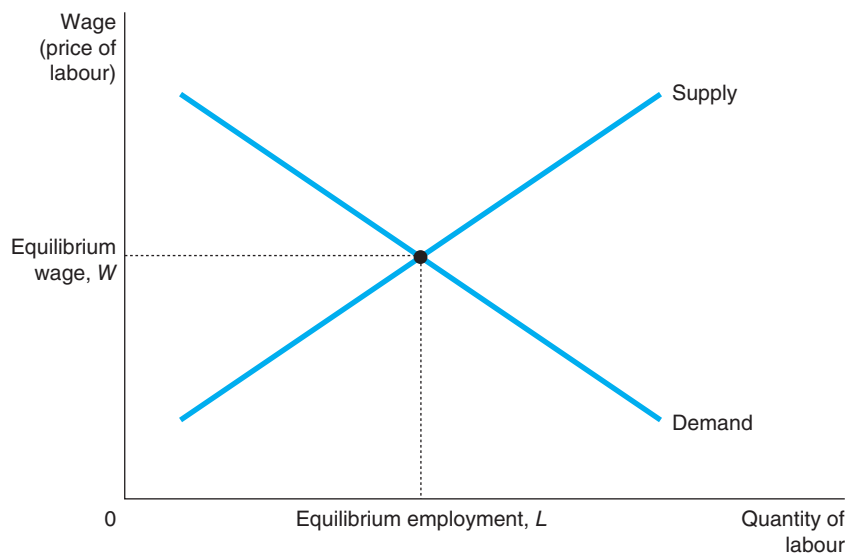
Figure 17.5 shows the labour market in equilibrium. The wage and the quantity of labour have adjusted to balance supply and demand. When the market is in this equilibrium, each firm has bought as much labour as it finds profitable at the equilibrium wage. That is, each firm has followed the rule for profit maximization: it has hired workers until the value of the marginal product equals the wage. Hence, the wage must equal the value of the marginal product of labour once it has brought supply and demand into equilibrium.

This brings us to an important lesson: any event that changes the supply or demand for labour must change the equilibrium wage and the value of the marginal product by the same amount, because these must always be equal. To see how this works, let's consider some events that shift these curves.

FIGURE 17.5

Equilibrium in a Labour Market

Like all prices, the price of labour (the wage) depends on supply and demand. Because the demand curve reflects the value of the marginal product of labour, in equilibrium workers receive the value of their marginal contribution to the production of goods and services.



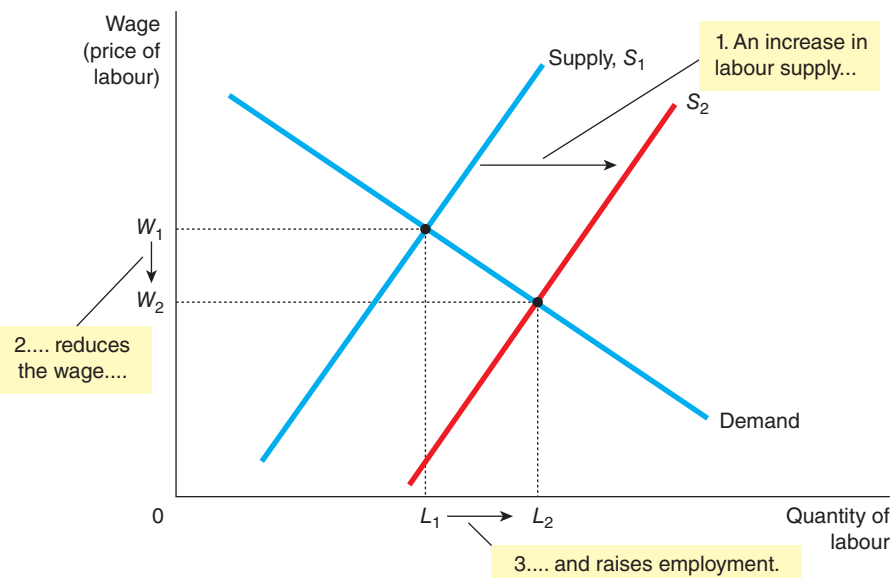
Shifts in Labour Supply

Suppose that immigration increases the number of workers willing to pick apples. As Figure 17.6 shows, the supply of labour shifts to the right from S_1 to S_2 . At the initial wage W_1 , the quantity of labour supplied now exceeds the quantity demanded. This surplus of labour puts downward pressure on the wage of apple pickers, and the fall in the wage from W_1 to W_2 in turn makes it profitable for firms to hire more workers. As the number of workers employed in each apple orchard rises, the marginal product of a worker falls, and so does the value of the marginal product. In the new equilibrium, both the wage and the value of the marginal product of labour are lower than they were before the influx of new workers.

FIGURE 17.6

A Shift in Labour Supply

When labour supply increases from S_1 to S_2 , perhaps because of immigration of new workers, the equilibrium wage falls from W_1 to W_2 . At this lower wage, firms hire more labour, so employment rises from L_1 to L_2 . The change in the wage reflects a change in the value of the marginal product of labour: with more workers, the added output from an extra worker is smaller.



Shifts in Labour Demand

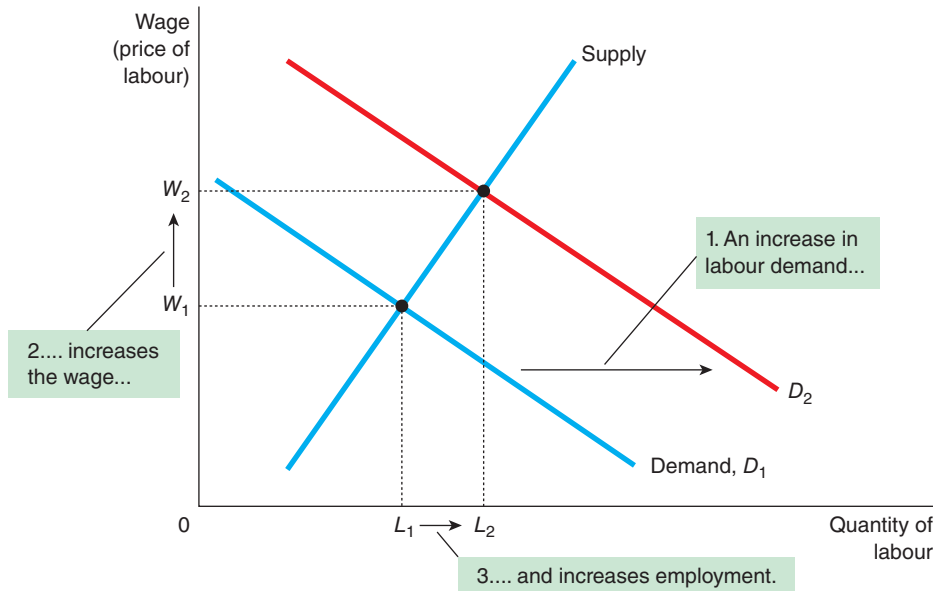
Now suppose that an increase in the popularity of apples causes their price to rise. This price increase does not change the marginal product of labour for any given number of workers, but it does raise the *value* of the marginal product. With a higher price of apples, hiring more apple pickers is now profitable. As Figure 17.7 shows, when the demand for labour shifts to the right from D_1 to D_2 , the equilibrium wage rises from W_1 to W_2 , and equilibrium employment rises from L_1 to L_2 . Once again, the wage and the value of the marginal product of labour move together.

This analysis shows that prosperity for firms in an industry is often linked to prosperity for workers in that industry. When the price of apples rises, apple producers make greater profit and apple pickers earn higher wages. When the price of apples falls, apple producers earn smaller profit and apple pickers earn lower wages. This lesson is well known to workers in industries with highly volatile prices. Workers in oil fields, for instance, know from experience that their earnings are closely linked to the world price of crude oil.

In competitive labour markets, therefore, labour supply and labour demand together determine the equilibrium wage, and shifts in the supply or demand curve for labour cause the equilibrium wage to change. At the same time, profit maximization by the firms that demand labour ensures that the equilibrium wage always equals the value of the marginal product of labour.

FIGURE 17.7**A Shift in Labour Demand**

When labour demand increases from D_1 to D_2 , perhaps because of an increase in the price of the firms' output, the equilibrium wage rises from W_1 to W_2 , and employment rises from L_1 to L_2 . Again, the change in the wage reflects a change in the value of the marginal product of labour: with a higher output price, the added output from an extra worker is more valuable.



Monopsony

So far, we have built our analysis of the labour market with the tools of supply and demand and assumed that the labour market was competitive. That is, we assumed that there were many buyers of labour and many sellers of labour, so each buyer or seller had a negligible effect on the wage.

Yet imagine the labour market in a small town dominated by a single large employer. That employer can exert a large influence on the going wage, and it may well use that market power to alter the outcome. Such a market in which there is a single buyer is called a **monopsony**.

monopsony a market with a single buyer

A monopsony is in many ways similar to a monopoly. We have seen that a monopoly firm produces less of the good than would a competitive firm: by reducing the quantity offered for sale, the monopoly firm moves along the product's demand curve, raising the price and also its profits. Similarly, a monopsony firm in a labour market hires fewer workers than would a competitive firm: by reducing the number of jobs available, the monopsony firm moves along the labour supply curve, reducing the wage it pays and raising its profits. Thus, both monopolists and monopsonists reduce economic activity in a market below the socially optimal level. In both cases, the existence of market power distorts the outcome and causes deadweight losses.

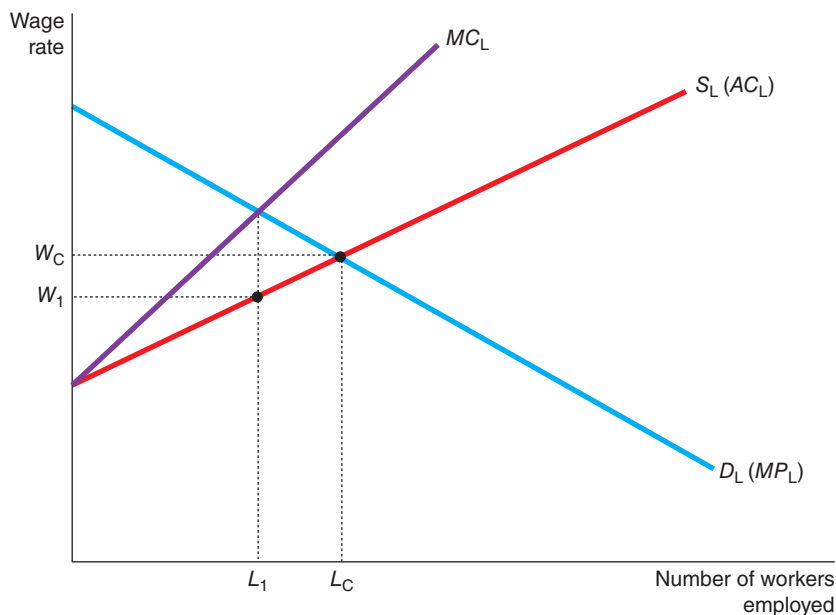
Figure 17.8 illustrates the situation. In a competitive market the employer would hire L_1 workers at a wage rate of W_1 where the demand curve for labour intersects the supply curve for labour. In a situation where an employer has monopsony power, the employer will take into account that fact that the supply curve for labour represents the average cost of labour. At the competitive wage W_1 , the number of people employed is L_1 and so the total cost of employing L_1 workers is the wage rate multiplied by the number of workers ($W_1 \times L_1$). The average cost of employing labour is thus the total cost divided by the number of workers, which is the wage rate W_1 .

A monopsonist will look to equate the marginal cost of labour with the marginal product. The MC of labour will be higher than the average cost of labour if the average cost of labour is rising. Each additional unit of labour employed must be higher than the average to keep the average rising. The MC of labour is shown by the curve MC_L . If the monopsonist sets the employment level where the MC of labour equals the MP of labour, the number of workers employed will be L_2 , lower than the competitive number of workers employed. If the employer wishes to employ L_2 amount of labour then the wage rate workers will accept will be given by the supply curve of labour at W_2 . The wage rate in a monopsony situation will therefore be lower than the competitive wage rate.

FIGURE 17.8

The Wage Rate and Employment Level in a Monopsony

An employer with monopsony will set the number of workers employed where the marginal cost of labour equals the marginal product. The monopsonist will employ a lower number of workers than the competitive market outcome and at a lower wage rate.



In the real world, monopsonies are rare although on a small scale, a number of towns in parts of Europe may be highly dependent on a major employer – a motor vehicle manufacturer, a steel works or chocolate manufacturer, for example. In such situations, the analysis may have to be amended to take into consideration the effect that monopoly power of the employer has on the local labour market. In most labour markets, however, workers have many possible employers, and firms compete with one another to attract workers. In this case, the model of supply and demand is the best one to use.

SELF TEST How does emigration of workers from a country affect labour supply, labour demand, the marginal product of labour and the equilibrium wage?

WAGE DIFFERENTIALS

In most economies there are significant differences in the earnings of workers. Why do earnings vary so much from person to person? We have seen that wages are governed by labour supply and labour demand. Labour demand, in turn, reflects the marginal productivity of labour. In equilibrium, each worker is paid the value of his or her marginal contribution to the economy's production of goods and services.

This theory of the labour market, though widely accepted by economists, is only the beginning of the story. To understand the wide variation in earnings that we observe, we must go beyond this general framework and examine more precisely what determines the supply and demand for different types of labour.

Workers differ from one another in many ways. Jobs also have differing characteristics – both in terms of the wage they pay and in terms of their non-monetary attributes. In this section we consider how the characteristics of workers and jobs affect labour supply, labour demand and equilibrium wages.

Compensating Differentials

When a worker is deciding whether to take a job, the wage is only one of many job attributes that the worker takes into account. Some jobs require few skills, and are ‘easy’ and safe; others might require considerable skill, experience and may be very dull, whilst others can be very dangerous. The ‘better’ the job as gauged by these non-monetary characteristics, the more people there are who are willing (and able) to do the job at any given wage. In other words, the supply of labour for jobs requiring few skills or no experience is greater than the supply of labour for highly skilled and dangerous jobs. As a result, ‘good’ jobs will tend to have lower equilibrium wages than ‘bad’ jobs.

For example, imagine you are looking for a summer job at a seaside holiday resort. Two kinds of jobs are available. You can take a job hiring out sunbeds on the beach, or you can take one as a refuse collector. The sunbed attendants take leisurely strolls along the beach during the day and check to make sure the holidaymakers have paid for the hire of their sunbeds. The refuse collectors wake up before dawn to drive dirty, noisy lorries around town to pick up rubbish from households and firms. Which job would you want? Most people would prefer the beach job if the wages were the same. To induce people to become refuse collectors, the town has to offer higher wages to refuse collectors than to sunbed attendants.

Economists use the term **compensating differential** to refer to a difference in wages that arises from non-monetary characteristics of different jobs. Compensating differentials are prevalent in the economy. Here are some examples:

compensating differential a difference in wages that arises to offset the non-monetary characteristics of different jobs

- Workers, who maintain and repair major roads such as motorways, are paid more than other public sector workers who repair roads in towns and cities. This is because the danger level of working on major roads is much higher, not to mention the fact that they often have to work unsociable hours (when drivers are not using the motorways).
- Workers who work night shifts at factories and in other forms of employment such as 24-hour retail outlets are paid more than similar workers who work day shifts. The higher wage compensates them for having to work at night and sleep during the day, a lifestyle that most people find undesirable (and disorientating!).
- University lecturers and professors are on average paid less than lawyers and doctors, who have similar levels of education. Lecturers’ lower wages are offset by the great intellectual and personal satisfaction that their jobs offer.

Human Capital

Human capital is the accumulation of investments in people. The most important type of human capital is education. Like all forms of capital, education represents an expenditure of resources at one point in time to raise productivity in the future. But, unlike an investment in other forms of capital, an investment in education is tied to a specific person, and this linkage is what makes it human capital.

human capital the accumulation of investments in people, such as education and on-the-job training

Not surprisingly, workers with more human capital on average earn more than those with less human capital. University graduates in Europe and North America, for example, earn almost twice as much as those workers who end their education after secondary school. This large difference tends to be even larger in less developed countries, where educated workers are in scarce supply.

It is easy to see why education raises wages from the perspective of supply and demand. Firms are willing to pay more for the highly educated because highly educated workers have higher marginal products. Workers are willing to pay the cost of becoming educated only if there is a reward for doing so. In essence, the difference in wages between highly educated workers and less educated workers may be considered a compensating differential for the cost of becoming educated.

Ability, Effort and Chance

Why do football players in the top European leagues such as the English Premiership or the Spanish *La Liga* get paid more than those in the minor leagues? Certainly, the higher wage is not a compensating differential. Playing in the major leagues is associated with more glamour and prestige as well as the, often considerably, higher wages. The major leagues do not necessarily require more years of schooling or more experience. To a large extent, players in the major leagues earn more just because they have greater natural ability.

Natural ability is important for workers in all occupations. Because of heredity and upbringing, people differ in their physical and mental attributes. Some people have physical and mental strength, whereas others have less of both. Some people are able to solve complex problems, others less so. Some people are outgoing, others awkward in social situations. These and many other personal characteristics determine how productive workers are and, therefore, play a role in determining the wages they earn.

Closely related to ability is effort. Some people are prepared to put long hours and considerable effort into their work whereas others are content to do what they are required to do and no more. We should not be surprised to find that those who put in more effort may be more productive and earn higher wages. To some extent, firms reward workers directly by paying people on the basis of what they produce. Salespeople, for instance, are often paid based on a percentage of the sales they make. At other times, greater effort is rewarded less directly in the form of a higher annual salary or a bonus.

Chance also plays a role in determining wages. If a person attended college to learn how to repair analogue devices and then found this skill made obsolete by the developments in digital technology, he or she would end up earning a low wage compared to others with similar years of training. The low wage of this worker is due to chance – a phenomenon that economists recognize but do not shed much light on.

How important are ability, effort and chance in determining wages? It is hard to say, because ability, effort and chance are hard to measure. But indirect evidence suggests that they are very important. When labour economists study wages, they relate a worker's wage to those variables that can be measured – years of schooling, years of experience, age and job characteristics. Although all of these measured variables affect a worker's wage as theory predicts, they account for less than half of the variation in wages in our economy. Because so much of the variation in wages is left unexplained, omitted variables, including ability, effort and chance, must play an important role.

An Alternative View of Education: Signalling

Although the view of the human capital theory is widely accepted, some economists have proposed an alternative theory, which emphasizes that firms use educational attainment as a way of sorting between high-ability and low-ability workers. According to this alternative view, when people earn a university degree, for instance, they do not become more productive, but they do *signal* their high ability to prospective employers. Because it is easier for high-ability people to earn a university degree than it is for low-ability people, more high-ability people get degrees. As a result, it is rational for firms to interpret a degree as a signal of ability.

The signalling theory of education is similar to the signalling theory of advertising. In the signalling theory of advertising, as we have seen, the advertisement itself contains no real information, but the firm signals the quality of its product to consumers by its willingness to spend money on advertising. In the signalling theory of education, schooling has no real productivity benefit, but the worker signals his innate productivity to employers by his willingness to spend years at school. In both cases, an action is being taken not for its intrinsic benefit but because the willingness to take that action conveys private information to someone observing it.

Thus, we now have two views of education: the human capital theory and the signalling theory. Both views can explain why more educated workers tend to earn more than less educated workers. According to the human capital view, education makes workers more productive; according to the signalling view, education is correlated with natural ability. But the two views have radically different predictions for the effects of policies that aim to increase educational attainment. According to the human capital view, increasing educational levels for all workers would raise all workers' productivity and thereby their wages. According to the signalling view, education does not enhance productivity, so raising all workers' educational levels would not affect wages.

Most likely, the truth lies somewhere between these two extremes. The benefits to education are probably a combination of the productivity enhancing effects of human capital and the productivity revealing effects of signalling. The open question is the relative size of these two effects.

The Superstar Phenomenon

Although most actors earn little and often take jobs as waiters to support themselves, some earns millions of pounds for taking part in a production. Similarly, while most people who play tennis do it for free as a hobby, some earns millions on the professional circuit. These individuals are superstars in their fields, and often their great public appeal is reflected in astronomical incomes.

Why do superstars earn so much? It is not surprising that there are differences in incomes within occupations. Good carpenters earn more than mediocre carpenters, and good plumbers earn more than mediocre plumbers. People vary in ability and effort, and these differences lead to differences in income. Yet the best carpenters and plumbers do not earn the many millions that are common among the best actors and athletes. What explains the difference? To understand the tremendous incomes of some individuals, we must examine the special features of the markets in which they sell their services. Superstars arise in markets that have two characteristics:

- Every customer in the market wants to enjoy the good supplied by the best producer.
- The good is produced with a technology that makes it possible for the best producer to supply every customer at low cost.

If Daniel Craig, for example, is the best actor around, then everyone will want to see his next film; seeing twice as many films by an actor half as good is not a good substitute. Moreover, it is *possible* for everyone to enjoy Daniel Craig's acting skills. Because it is easy to make multiple copies of a film, Daniel Craig can provide his service to millions of people simultaneously. According to Box Office Mojo, Daniel Craig's lifetime gross total box office takings amount to €1,013,209,688, an average of €44,052,595 per movie; that's a considerable value of marginal product to the film producers! Similarly, because sport is broadcast on television, millions of fans can enjoy the athletic skills of footballers, rugby players, athletes, basketball players and so on.

We can now see why there are no superstar carpenters and plumbers. Other things equal, everyone prefers to employ the best carpenter, but a carpenter, unlike a film actor, can provide his services to only a limited number of customers. Although the best carpenter will be able to command a somewhat higher wage than the average carpenter, the average carpenter will still be able to earn a good living.

Above-Equilibrium Wages: Minimum Wage Laws, Unions and Efficiency Wages

For some workers, wages are set above the level that brings supply and demand into equilibrium. Let's consider three reasons why this might be so.

Minimum Wage Laws Minimum wage laws are an example of a price floor. **Minimum wage** laws dictate the lowest price for labour that any employer may pay. Panel (a) of Figure 17.9 shows the labour market with the wage adjusting to balance labour supply and labour demand.

minimum wage the lowest price an employer may legally pay to a worker

Panel (b) of Figure 17.9 shows the labour market with a minimum wage. If the minimum wage is above the equilibrium level, as it is here, the quantity of labour supplied exceeds the quantity demanded. The result is unemployment. Thus, the minimum wage raises the incomes of those workers who have jobs, but it lowers the incomes of those workers who cannot find jobs.

To understand fully the minimum wage, keep in mind that the economy contains not a single labour market, but many labour markets for different types of workers. The impact of the minimum wage depends on the skill and experience of the worker. Workers with high skills and much experience are not affected, because their equilibrium wages are well above the minimum. For these workers, the minimum wage is not binding. One would therefore expect a diagram such as that in panel (b) of Figure 17.9, where the minimum wage is above the equilibrium wage and unemployment results, to apply primarily to the market for low-skilled and teenage labour. Note, however, that the *extent* of the unemployment that results depends upon the elasticities of the supply and demand for labour. In panel (c) of Figure 17.9 we have redrawn the diagram with a more elastic demand curve for labour and we can see that this results in a higher level of unemployment. It is often argued that the demand for unskilled labour is in fact likely to be highly elastic with respect to the price of labour because employers of unskilled labour, such as fast food restaurants, usually face highly price-elastic demand curves for their own product and so cannot easily pass on wage rises in the form of higher prices without seeing their revenue fall.

This is only true, however, if one firm raises its price while others do not. If all fast food companies are forced to raise prices slightly in order to pay the minimum wage to their staff, this may result in a much smaller fall in the demand for the output (e.g. hamburgers) of any one firm. If this is the case, then the imposition of a statutory minimum wage may actually lead to a rightward shift in the segment of the labour demand curve at or above the statutory minimum wage: a firm is able to pay the higher wage without drastically reducing its labour demand because it can pass on the higher wage costs by charging a higher price for its product, safe in the knowledge that other firms in the industry will have to do the same and hence that it will not suffer a dramatic fall in demand for its output. In this case – as in panel (d) of Figure 17.9 – although there is an increase in unemployment relative to the case with no minimum wage, this is mainly because the supply of labour is higher with the minimum wage imposed. This is because some workers will be attracted by the higher wage to enter the labour market – second earners, for example, or young people who otherwise would have stayed in full-time education.

Advocates of minimum wage laws view the policy as one way to raise the income of the working poor. They correctly point out that workers who earn the minimum wage can afford only a meagre standard of living. They admit that it may have some adverse effects, including a possible rise in unemployment, but they believe that these effects are small and that, all things considered, a higher minimum wage makes the poor better off. In other words they argue that the value of the benefits of a minimum wage are greater than the value of the costs and so such a policy is worth putting into practice.

Opponents of the minimum wage contend that it is not the best way to combat poverty since it affects only the income of those in employment and may raise unemployment, and because not all minimum wage workers are heads of households trying to help their families escape poverty – some may be second earners or even third earners in relatively well-off households. To decide whether this argument is more powerful than the arguments of the advocates, economists will try and find ways to measure the size of the contrasting effects so that an informed decision can be made with regard to the value of the benefits versus the costs. This is often harder than it may at first appear but is a crucial part of an economist's work.

In short, the effects of the minimum wage on the labour market are complicated. It is one of those cases where there are no clear-cut answers: it all depends on what assumptions you make. All our brief economic analysis of the minimum wage has shown is that the issue is very complex.

The Market Power of Labour Unions A second reason that wages might rise above their equilibrium level is the market power of labour unions. A **union** is a worker association that bargains with employers over wages and working conditions. Unions often raise wages above the level that would prevail without a union,

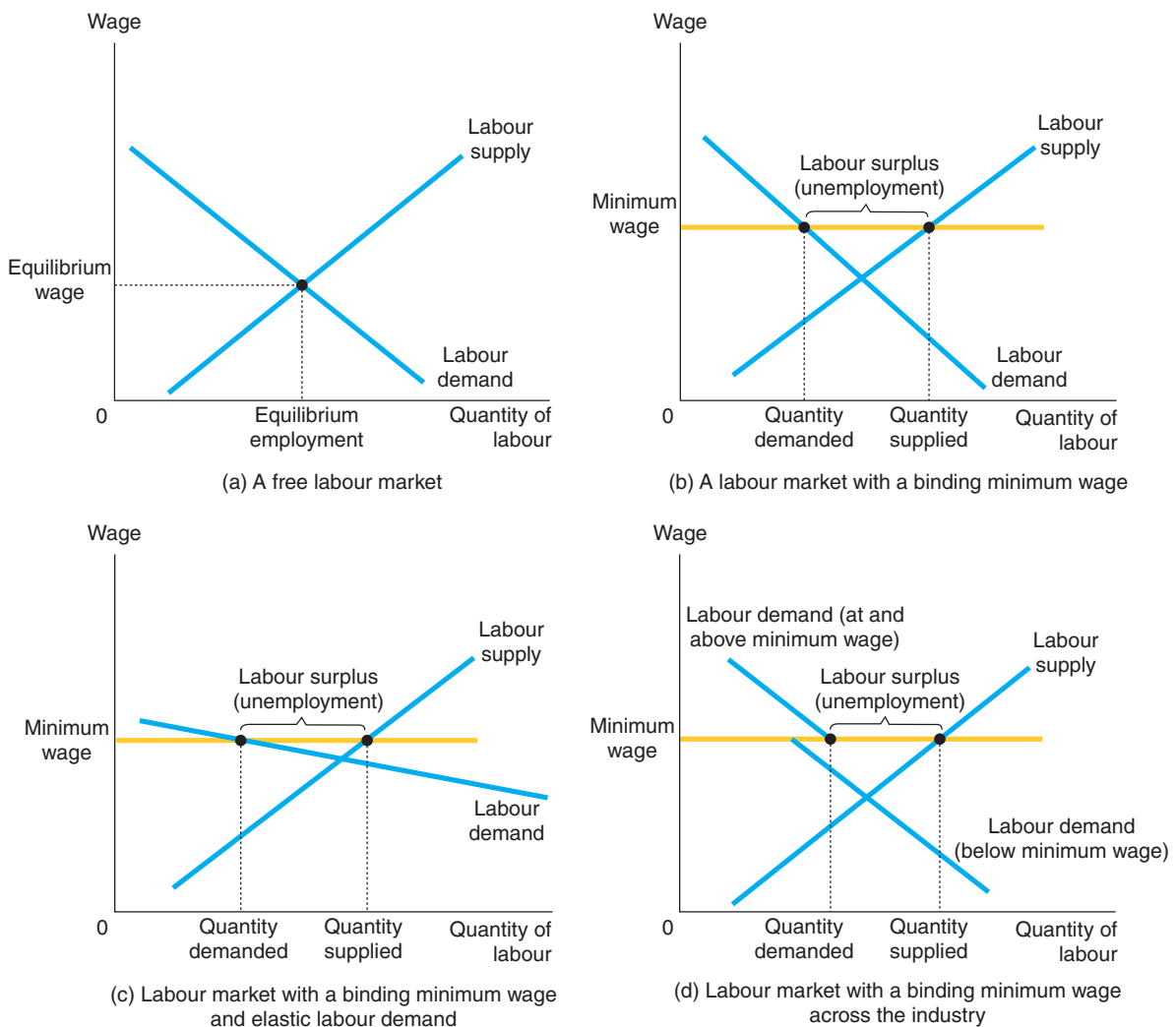
perhaps because they can threaten to withhold labour from the firm by calling a **strike**. Studies suggest that union workers earn about 10 to 20 per cent more than similar non-union workers.

strike the organized withdrawal of labour from a firm by a union

FIGURE 17.9

How the Minimum Wage Affects the Labour Market

Panel (a) shows a labour market in which the wage adjusts to balance labour supply and demand. Panel (b) shows the impact of a binding minimum wage. Because the minimum wage is a price floor, it causes a surplus: the quantity of labour supplied exceeds the quantity demanded. The result is unemployment. Panel (c) shows that the more elastic labour demand is, the higher will be ensuing unemployment. In panel (d), because the minimum wage is binding across the whole industry, firms are able to pass a higher proportion of the wage costs on to higher prices without a drastic fall in demand for output, and so the labour demand curve for an individual firm actually shifts to the right at or above the minimum wage, so that the impact on employment is much less.



Efficiency Wages A third reason for above-equilibrium wages is suggested by the theory of **efficiency wages**. This theory holds that a firm can find it profitable to pay high wages because doing so increases the productivity of its workers. In particular, high wages may reduce worker turnover (hiring and training)

new workers is an expensive business), increase worker effort, and raise the quality of workers who apply for jobs at the firm. In addition, a firm may feel it has to offer high wages in order to attract and keep the best people – this has been an argument put forward by the banking sector in response to plans by governments in Europe to tax bankers' earnings in the wake of the financial crisis. If this theory is correct, then some firms may choose to pay their workers more than they would normally earn.

efficiency wages above-equilibrium wages paid by firms in order to increase worker productivity

Above-equilibrium wages, whether caused by minimum wage laws, unions or efficiency wages, have similar effects on the labour market. In particular, pushing a wage above the equilibrium level raises the quantity of labour supplied and reduces the quantity of labour demanded. The result is a surplus of labour, or unemployment. The study of unemployment and the public policies aimed to deal with it is usually considered a topic within macroeconomics, so it goes beyond the scope of this chapter. But it would be a mistake to ignore these issues completely when analysing earnings. Although most wage differences can be understood while maintaining the assumption of equilibrium in the labour market, above-equilibrium wages play a role in some cases.

SELF TEST Define *compensating differential* and give an example. • Give two reasons why more educated workers earn more than less educated workers.

THE ECONOMICS OF DISCRIMINATION

Another source of differences in wages is discrimination. **Discrimination** occurs when the marketplace offers different opportunities to similar individuals who differ only by race, ethnic group, sex, age or other personal characteristics. Discrimination reflects some people's prejudice against certain groups in society. Although discrimination is an emotionally charged topic that often generates heated debate, economists try to study the topic objectively in order to separate myth from reality.

discrimination the offering of different opportunities to similar individuals who differ only by race, ethnic group, sex, age or other personal characteristics

Measuring Labour Market Discrimination

How much does discrimination in labour markets affect the earnings of different groups of workers? This question is important, but answering it is not easy. There is no doubt that different groups of workers earn substantially different wages.

A report published in 2010 by the National Equality Panel (Hills, J. *et al.* (2010). *An Anatomy of Economic Inequality in the UK* showed that inequalities in incomes and earnings are 'high' in Britain compared with other industrialized countries, and considering figures from 30 years ago. Women, for example, earn 21 per cent less in terms of median hourly pay for all employees and 13 per cent less than men for those working full time. The Report also found that people from non-white ethnic backgrounds are paid less than white British men. Pakistani and Bangladeshi Muslim men and Black African Christian men earn between 13 and 21 per cent less than white British Christian men. Chinese men are one of the highest paid groups in Britain but they are paid 11 per cent less than would be expected given their qualifications. Women from nearly all ethno-religious backgrounds have pay which is anything between a quarter and a third less than their white British Christian male counterparts, even with the same qualifications, age and in the same occupation, according to the Report.

In Germany, women earned around 24 per cent less than men in 2009 compared with the European Union average of 18 per cent. Of the 28 member countries of the EU, only in Estonia, the Czech Republic,

Austria and the Netherlands is there a higher gender gap in pay in favour of men. Pay differentials between men and women across Europe have hardly changed over the past 15 years according to the European Commission.

Taken at face value, these differentials look like evidence that UK and European employers discriminate against those from ethnic minorities and women. Yet there is a potential problem with this inference. Even in a labour market free of discrimination, different people have different wages. People differ in the amount of human capital they have and in the kinds of work they are able and willing to do. The wage differences we observe in the economy are, to some extent, attributable to the determinants of equilibrium wages we discussed in the preceding section. Simply observing differences in wages among broad groups – whites and blacks, men and women – does not prove that employers discriminate.

Consider, for example, the role of human capital. Whether an individual has a degree can account for some of these differences and as we have seen above, the type of degree can also have an impact. Those with a degree in sciences may earn more than those who have degrees from the humanities and arts. Human capital may also be more important in explaining wage differentials than measures of years of schooling suggest. The quality of education might affect the quality of human capital. The quality of education both in schools and at higher education can also be measured by expenditure, class size, ratio of teachers to pupils and so on. If we could measure the quality as well as the quantity of education, the differences in human capital among these groups would seem even larger.

The Federal Bureau of Statistics in Germany points out that gender pay differences in Germany may be due to a number of factors including differences in educational attainment, a high proportion of women work in part-time occupations (69 per cent of employed mothers in 2009) and the type of employment they go into; many of the jobs women enter tend to be low-skill, low-paid jobs.

Human capital acquired in the form of job experience can also help explain wage differences. In particular, women tend to have less job experience on average than men. One reason is that female labour force participation has increased in industrialized economies over the past several decades. Because of this historic change, in both Europe and North America, the average female worker today is younger than the average male worker. In addition, women are more likely to interrupt their careers to raise children. For both reasons, the experience of the average female worker is less than the experience of the average male worker.

Men and women do not always choose the same type of work, and this fact may help explain some of the earnings differential between men and women. For example, women are more likely to be personal assistants or receptionists, and men are more likely to be lorry drivers. (Some people would argue that women get 'ushered' into these types of jobs because of stereotypes, but that is not the point here.) The relative wages of personal assistants, receptionists and lorry drivers depend in part on the working conditions of each job. Because these non-monetary aspects are hard to measure, it is difficult to gauge the practical importance of compensating differentials in explaining the wage differences that we observe.

In the end, the study of wage differences among groups does not establish any clear conclusion about the prevalence of discrimination in labour markets. Most economists believe that some of the observed wage differentials are attributable to discrimination, but there is no consensus about how much. The only conclusion about which economists are in consensus is a negative one: because the differences in average wages among groups in part reflect differences in human capital and job characteristics, they do not by themselves say anything about how much discrimination there is in the labour market.

Of course, differences in human capital among groups of workers may themselves reflect discrimination. The less rigorous curriculums historically offered to female students, for instance, can be considered a discriminatory practice. Similarly, the inferior schools available to those who live in poorer areas could be traced to prejudice on the part of local governments. But this kind of discrimination occurs long before the worker enters the labour market. In this case, the disease is political, even if the symptom is economic.

Discrimination by Employers

Let's now turn from measurement to the economic forces that lie behind discrimination in labour markets. If one group in society receives a lower wage than another group, even after controlling for human capital and job characteristics, who is to blame for this differential?

The answer is not obvious. It might seem natural to blame employers for discriminatory wage differences. After all, employers make the hiring decisions that determine labour demand and wages. If some groups of workers earn lower wages than they should, then it seems that employers are responsible.

Yet many economists are sceptical of this easy answer. They believe that competitive market economies provide a natural antidote to employer discrimination. That antidote is called the profit motive.

Imagine an economy in which workers are differentiated by their hair colour. Blondes and brunettes have the same skills, experience and work ethic. Yet, because of discrimination, employers prefer not to hire workers with blonde hair. Thus, the demand for blondes is lower than it otherwise would be. As a result, blondes earn a lower wage than brunettes.

How long can this wage differential persist? In this economy, there is an easy way for a firm to beat out its competitors: it can hire blonde workers. By hiring blondes, a firm pays lower wages and thus has lower costs than firms that hire brunettes. Over time, more and more 'blonde' firms enter the market to take advantage of this cost advantage. The existing 'brunette' firms have higher costs and, therefore, begin to lose money when faced with the new competitors. These losses induce the brunette firms to go out of business. Eventually, the entry of blonde firms and the exit of brunette firms cause the demand for blonde workers to rise and the demand for brunette workers to fall. This process continues until the wage differential disappears.

Put simply, business owners who care only about making money are at an advantage when competing against those who also care about discriminating. As a result, firms that do not discriminate tend to replace those that do. In this way, competitive markets have a natural remedy for employer discrimination.

Discrimination by Customers and Governments

Although the profit motive is a strong force acting to eliminate discriminatory wage differentials, there are limits to its corrective abilities. Here we consider two of the most important limits: *customer preferences* and *government policies*.

To see how customer preferences for discrimination can affect wages, consider again our imaginary economy with blondes and brunettes. Suppose that restaurant owners discriminate against blondes when hiring waiters. As a result, blonde waiters earn lower wages than brunette waiters. In this case, a restaurant could open up with blonde waiters and charge lower prices. If customers only cared about the quality and price of their meals, the discriminatory firms would be driven out of business, and the wage differential would disappear.

On the other hand, it is possible that customers prefer being served by brunette waiters. If this preference for discrimination is strong, the entry of blonde restaurants need not succeed in eliminating the wage differential between brunettes and blondes. That is, if customers have discriminatory preferences, a competitive market is consistent with a discriminatory wage differential. An economy with such discrimination would contain two types of restaurants. Blonde restaurants hire blondes, have lower costs and charge lower prices. Brunette restaurants hire brunettes, have higher costs and charge higher prices. Customers who did not care about the hair colour of their waiters would be attracted to the lower prices at the blonde restaurants. Bigoted customers would go to the brunette restaurants. They would pay for their discriminatory preference in the form of higher prices.

Another way for discrimination to persist in competitive markets is for the government to mandate discriminatory practices. If, for instance, the government passed a law stating that blondes could wash dishes in restaurants but could not work as waiters, then a wage differential could persist in a competitive market. Before South Africa abandoned its system of Apartheid, blacks were prohibited from working in some jobs. Discriminatory governments pass such laws to suppress the normal equalizing force of free and competitive markets.

To sum up: competitive markets contain a natural remedy for employer discrimination. The entry into the market of firms that care only about profit tends to eliminate discriminatory wage differentials. These wage differentials persist in competitive markets only when customers are willing to pay to maintain the discriminatory practice or when the government mandates it.

Becker's 'Employer Taste' Model

One important piece of research into the economics of discrimination is from Nobel Prize winner Gary Becker from the University of Chicago, who in 1971 revised his earlier 1957 work on this area. The basis of the employer taste model is that (for whatever reason) some employees will resist working with other

employees, possibly because of gender or race. People may, therefore, have a 'taste' for only working with certain groups of people. Those outside this accepted group may end up being disadvantaged as a result.

Assume that a UK firm, which grows asparagus, hires workers to cut the spears. It has a choice of employing locals or migrant workers. Local people have a prejudice against migrant workers for some reason. Our analysis of a competitive firm assumes that workers will be employed up to the point at which the wage equals the marginal revenue product of labour. Assume that both local and migrant workers have the same level of productivity. If the firm has to employ workers at a going wage (which is above the minimum wage) then it may choose not to employ workers from the disadvantaged group because of the preferences of its core workforce. If, however, the firm is able to pay workers from the disadvantaged group lower wages then it faces a trade-off. There is an incentive for it to increase profits by employing these 'disadvantaged' workers – the migrant workers from Europe. If migrant workers were prepared to work for the minimum wage then the firm could lower its costs and increase profit as a result.

A discriminatory firm might employ some migrant workers but would pay these workers a lower wage to avoid upsetting their local workers. This is the 'employer taste' model – discrimination will exist because employers do not employ labour from certain genders, race, etc. unless the workers are prepared to accept lower wages. This discrimination may continue whilst there is some limit to the competition in the labour market – in this case it might be that all firms are prepared to act in the same way.

However, if there were other asparagus farms in the area who were not discriminatory then a firm might choose to hire all workers at the minimum wage which would increase its overall profits. Such a firm would also employ more workers (remember that the lower the wage rate the more workers a firm is willing to employ). There could be an influx of migrant workers to the area who are willing to take advantage of the jobs available. These non-discriminatory firms could not only produce more output but at a lower wage cost per unit and so make more profit, possibly driving out the discriminatory firm from the industry.

In the UK, such a situation has manifested itself in recent years. The extension of membership of the EU in 2004 led to an increase in the number of migrant workers from countries such as Poland, Lithuania and the Czech Republic coming to Britain to find work. Many of these workers appeared willing to take on jobs that paid relatively low wages, such as cutting asparagus spears. In Cambridgeshire in the southeast of England, a large number found work on the farms in the region, picking and packing fruit and vegetables. In the town of Wisbech, for example, there are over 2,000 'local' people who are unemployed but there have been around 9,000 migrant workers who have secured jobs in the area – mostly in jobs where the wages are traditionally low.

The sensitivity of the situation in Wisbech, with some of the local unemployed blaming migrant workers for their lack of work, is difficult. Some employers have been accused of exploiting migrant labour by paying them low wages but some counter that they are paying at least the minimum wage and that they find migrant workers not only willing to work for lower pay but also that their productivity levels are relatively high compared with some 'local' labour. In this case not only are migrant workers prepared to work for lower wages but their marginal product is higher at each price (wage). Some farmers claim that 'local' workers are not prepared to do the sort of work that is available and believe that it is too low paid. It seems that regardless of discrimination, employers are more concerned with getting value for money from their employees and are prepared to put profit before discrimination.

SELF TEST Why is it hard to establish whether a group of workers is being discriminated against? ● Explain how profit-maximizing firms tend to eliminate discriminatory wage differentials. ● How might a discriminatory wage differential persist?

THE OTHER FACTORS OF PRODUCTION: LAND AND CAPITAL

We have seen how firms decide how much labour to hire and how these decisions determine workers' wages. At the same time that firms are hiring workers, they are also deciding about other inputs to production. For example, our apple-producing firm might have to choose the size of its apple orchard (land) and the number of ladders to make available to its apple pickers, the baskets that are used to collect the

picked apples, trucks used to transport the apples the buildings used to store the apples and even the trees themselves (capital).

Equilibrium in the Markets for Land and Capital

What determines how much the owners of land and capital earn for their contribution to the production process? Before answering this question, we need to distinguish between two prices: the purchase price and the rental price. The *purchase price* of land or capital is the price a person pays to own that factor of production indefinitely. The *rental price* is the price a person pays to use that factor for a limited period of time. It is important to keep this distinction in mind because, as we will see, these prices are determined by somewhat different economic forces.

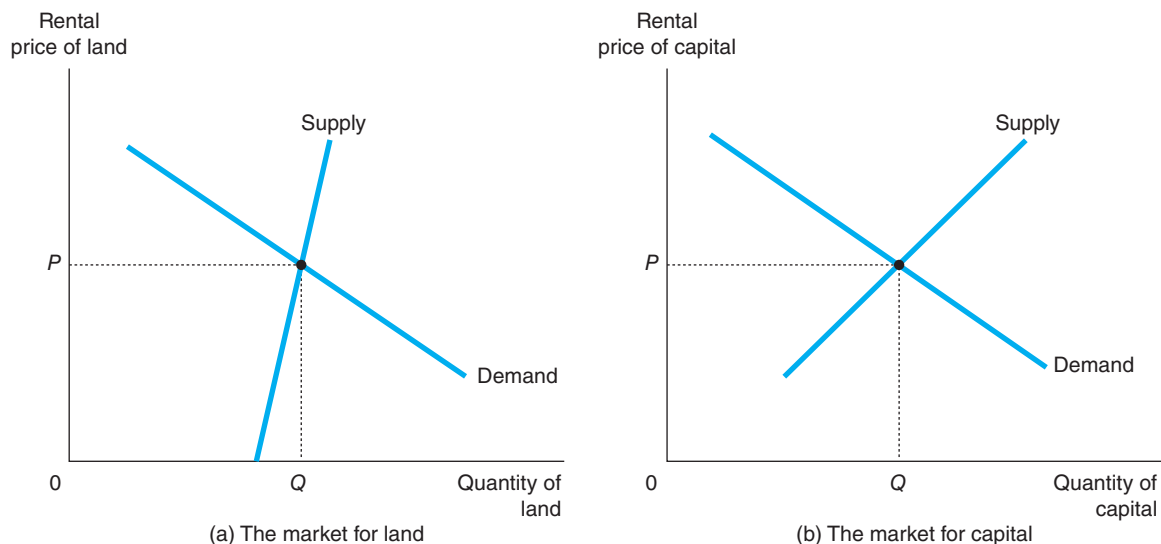
Having defined these terms, we can now apply the theory of factor demand that we developed for the labour market to the markets for land and capital. The wage is, after all, simply the rental price of labour. Therefore, much of what we have learned about wage determination applies also to the rental prices of land and capital. As Figure 17.10 illustrates, the rental price of land, shown in panel (a), and the rental price of capital, shown in panel (b), are determined by supply and demand. Moreover, the demand for land and capital is determined just like the demand for labour. That is, when our apple-producing firm is deciding how much land and how many ladders to rent, it follows the same logic as when deciding how many workers to hire. For both land and capital, the firm increases the quantity hired until the value of the factor's marginal product equals the factor's price. Thus, the demand curve for each factor reflects the marginal productivity of that factor.

We can now explain how much income goes to labour, how much goes to landowners and how much goes to the owners of capital. As long as the firms using the factors of production are competitive and profit-maximizing, each factor's rental price must equal the value of the marginal product for that factor: labour, land and capital each earn the value of their marginal contribution to the production process.

FIGURE 17.10

The Markets for Land and Capital

Supply and demand determine the compensation paid to the owners of land, as shown in panel (a), and the compensation paid to the owners of capital, as shown in panel (b). The demand for each factor, in turn, depends on the value of the marginal product of that factor.



Now consider the purchase price of land and capital. The rental price and the purchase price are obviously related: buyers are willing to pay more for a piece of land or capital if it produces a valuable stream of rental income. And, as we have just seen, the equilibrium rental income at any point in time equals the value of that factor's marginal product. Therefore, the equilibrium purchase price of a piece of land or capital depends on both the current value of the marginal product and the value of the marginal product expected to prevail in the future.

Linkages Among the Factors of Production

We have seen that the price paid to any factor of production – labour, land or capital – equals the value of the marginal product of that factor. The marginal product of any factor, in turn, depends on the quantity of that factor that is available. Because of diminishing marginal product, a factor in abundant supply has a low marginal product and thus a low price, and a factor in scarce supply has a high marginal product and a high price. As a result, when the supply of a factor falls, its equilibrium factor price rises.

When the supply of any factor changes, however, the effects are not limited to the market for that factor. In most situations, factors of production are used together in a way that makes the productivity of each factor dependent on the quantities of the other factors available to be used in the production process. As a result, a change in the supply of any one factor alters the earnings of all the factors.

For example, suppose one night lightning strikes the storehouse in which are kept the ladders that the apple pickers use to pick apples from the orchards, and many of the ladders are destroyed in the ensuing

FYI



What is Capital Income?

Labour income is an easy concept to understand: it is the wages and salaries that workers get from their employers. The income earned by capital, however, is less obvious.

In our analysis, we have been implicitly assuming that households own the economy's stock of capital – equipment, machinery, computers, warehouses and so forth – and rent it to the firms that use it. Capital income, in this case, is the rent that households receive for the use of their capital. This assumption simplified our analysis of how capital owners are compensated, but it is not entirely realistic. In fact, firms usually own the capital they use and, therefore, they receive the earnings from this capital.

These earnings from capital, however, eventually get paid to households. Some of the earnings are paid in the form of interest to

those households who have lent money to firms (anyone who has savings in a financial institution, who pays into a pension fund or an insurance policy is indirectly actually lending money to businesses!). Bondholders and bank depositors are two examples of recipients of interest. Thus, when you receive interest on your bank account, that income is part of the economy's capital income.

In addition, some of the earnings from capital are paid to households in the form of dividends. Dividends are payments by a firm to the firm's shareholders. A shareholder is a person who has bought a share in the ownership of the firm and, therefore, is entitled to share in the firm's profits. (This is usually called an equity or, quite simply, a share.)

A firm does not have to pay out all of its earnings to households in

the form of interest and dividends. Instead, it can retain some earnings within the firm and use these earnings to buy additional capital. Although these retained earnings do not get paid to the firm's shareholders, the shareholders benefit from them nonetheless. Because retained earnings increase the amount of capital the firm owns, they tend to increase future earnings and, thereby, the value of the firm's equities.

These institutional details are interesting and important, but they do not alter our conclusion about the income earned by the owners of capital. Capital is paid according to the value of its marginal product, regardless of whether this income gets transmitted to households in the form of interest or dividends or whether it is kept within firms as retained earnings.

fire. What happens to the earnings of the various factors of production? Most obviously, the supply of ladders falls and, therefore, the equilibrium rental price of ladders rises. Those owners who were lucky enough to avoid damage to their ladders now earn a higher return when they rent out their ladders to the firms that produce apples.

Yet the effects of this event do not stop at the ladder market. Because there are fewer ladders with which to work, the workers who pick apples have a smaller marginal product. Thus, the reduction in the supply of ladders reduces the demand for the labour of apple pickers, and this causes the equilibrium wage to fall.

This story shows a general lesson: an event that changes the supply of any factor of production can alter the earnings of all the factors. The change in earnings of any factor can be found by analysing the impact of the event on the value of the marginal product of that factor.

SELF TEST What determines the income of the owners of land and capital? ● How would an increase in the quantity of capital affect the incomes of those who already own capital? ● How would it affect the incomes of workers?

ECONOMIC RENT

Consider a professional footballer in a top European league. Some players in these leagues earn tens of thousands of euros a week. Assume a player earns €100,000 per week; if that player's wages were cut to €50,000, would he still be a professional footballer? What if his wages were cut to €20,000 a week, or €5,000 a week? (€5,000 a week is still €260,000 a year). At what point would the player make the decision to stop being a professional footballer and switch to doing something else instead?

Now consider a plot of land and a series of machines used by a business for manufacturing CDs. The demand for CDs is falling but the business could use the machines and factory to produce Blu-ray DVDs. At what point does the falling earnings from CDs lead to the business switching from CD production to Blu-ray DVD production? If the earnings from CD or Blu-ray production fall, at what point does the firm decide to switch the use of the land and capital away from that particular use and to another use altogether?

This is the subject of what is called economic rent. **Economic rent** is the amount a factor of production earns over and above its transfer earnings. **Transfer earnings** are, in turn, the minimum payment required to keep a factor of production in its current use. The transfer earnings of a factor, therefore, represents the opportunity cost of the factor being employed in its current use.

economic rent the amount a factor of production earns over and above its transfer earnings
transfer earnings the minimum payment required to keep a factor of production in its current use

Let us go back to our professional football example. Assume that a player is currently earning €200,000 per year. The player also happens to be a qualified chartered surveyor and assume that the average annual income for those in this profession is €88,400. Provided the player earns in excess of the amount he could earn as a surveyor, it is rational for the player to continue as a professional footballer. If the player's wages were cut to €100,000 a year then there would still be an incentive for the player to stay as a footballer. However, if the player's wages were cut to €85,000 then he could earn more from being a surveyor and it would make sense for him to transfer to that occupation.

The difference between the amount the factor earns and the transfer earnings is termed the economic rent. If the player earns €200,000 then the economic rent, in our example, is €111,600. This is the amount by which the player's earnings could fall before they transferred to an alternative employment.

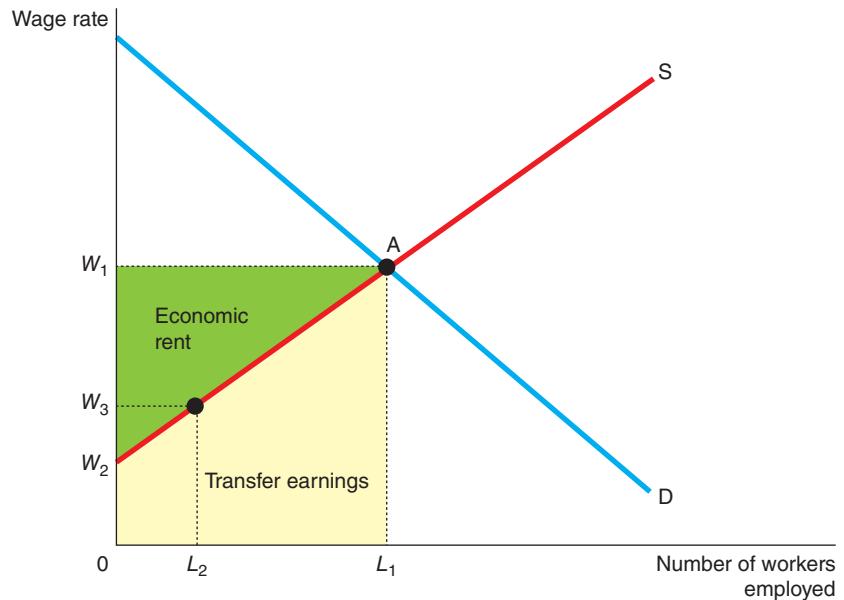
We can see the size of the economic rent in Figure 17.11. The wage rate is given by the intersection of the demand and supply curve at W_1 . If the wage rate was W_2 , the number of people willing to work in this industry would be zero (the vertical intercept of the supply curve). At wage rates higher than W_2 , more

workers are willing to offer their services. At a wage rate of W_3 , L_2 workers would be willing to offer their services. For the L_2 worker, the wage rate of W_3 is just sufficient to encourage them into that employment but for all the other workers up to L_2 the wage rate is higher than the amount they would be willing to earn to offer their services. A wage of W_3 , therefore, will yield some economic rent for these workers.

When L_1 workers are employed, the total economic rent is given by the area above the supply curve and the transfer earnings by the area below the supply curve.

FIGURE 17.11**Economic Rent**

At the market wage rate of W_1 , the economic rent is shown by the area above the supply curve and the transfer earnings by the area below the supply curve.



The principle of economic rent can be applied to all factors of production. It is important when thinking of land that you do not confuse the general use of the term *rent* with the economic definition. Economic rent has applications across a range of economic situations. In particular, it has been discussed with reference to taxation. If economic rent exists for any factor of production, the government could, in theory, tax a portion of that rent without affecting the employment of that factor in a particular use. A government might, therefore, debate how much of a banker's bonus to tax if there is an assumption that a large part of the earnings of bankers represents economic rent. Governments might consider taxing land and providing the tax does not push earnings below the transfer earnings then the land will continue to be used in its current form.

CONCLUSION

The theory developed in this chapter is called the *neoclassical theory of distribution*. According to the neoclassical theory, the amount paid to each factor of production depends on the supply and demand for that factor. The demand, in turn, depends on that particular factor's marginal productivity. In equilibrium, each factor of production earns the value of its marginal contribution to the production of goods and services.

The theory can be used to answer questions about why some workers are paid more than others. It is because some workers can produce a good of greater market value than can others and so the wages of workers reflect the market prices of the goods they produce. In competitive markets, workers earn a wage equal to the value of their marginal contribution to the production of goods and services. There are, however, many things that affect the value of the marginal product. Firms pay more for workers who are more talented, more diligent, more experienced and more educated because these workers are more productive. Firms pay less to those workers against whom customers discriminate because these workers contribute less to revenue.

IN THE NEWS



Labour Market Rigidities

In most countries, the labour market is not perfectly competitive. Various laws and regulations mean that different countries have different degrees of rigidity in the labour market which can affect the ability of firms to respond to changing economic conditions. Reducing some of these rigidities can be difficult and bring to the fore sharp divisions of opinion.

The Labour Market in France

Firms in France have to work under a number of labour laws which they claim limits their ability to be flexible and puts them at a competitive disadvantage compared to other countries across Europe. Once a firm employs more than 50 workers it faces additional costs in the form of stricter redundancy procedures and also has to accept that workers have greater representation rights. The marginal cost of employing the 50th worker, therefore, is higher than that of the 49th. Some firms seek to circumvent these regulations by not employing more than 49 workers or create two separate businesses, each doing the same thing but under different names, in order to avoid the additional costs.

For employees who are in work, the regulations offer them some protection; traditionally, fixed-term contracts over long periods have meant that workers have considerable security and the power of trade unions is such that employers find it difficult to introduce new working practices or to make excess labour redundant to maintain or improve productivity, especially in times of economic difficulty. The handing out of lengthy fixed-term contracts is on the decline with many firms turning to short-term contracts as a way of providing greater flexibility.

The change in approach, however, is slow. The country's labour code runs to in excess of 3,000 pages and many employers find that if they do try instituting redundancy programmes employees use the

code to challenge the decision in the courts and often find the legal system sympathetic to employees rather than employers.

The French government spent a considerable amount of time seeking agreement on reforms to the country's labour laws. In early 2013, a deal was finally agreed between the government, employers and unions to reform the labour market. Key to the agreement was the ability for firms to be more flexible about working time and pay during periods of economic slowdown. In other parts of Europe where economic difficulties have been as bad as in France, unemployment has been less than in France and one reason suggested has been the ability of firms to reduce the pay of workers or put them on a



An empty platform at the Gare de l'Est railway station in Paris during a strike by transportation workers. Such industrial action can paralyze the economy.

(Continued)

reduced-time week rather than making redundancies. Once the economy picks up the firm then goes back to normal working hours or adjusts pay. The cost of doing this is considerably less than going through redundancy procedures and then having to recruit employees with the necessary skills needed when business picks up. For the employer, the trade-off of a period of reduced time and/or pay is that he gets to keep his job and so does not have to endure the cost and pain of searching for a new job.

Questions

- 1 **What are the likely consequences of imperfections in the labour market compared to the model of a perfectly competitive labour market we have looked at in the first part of this chapter?**
- 2 **Why do you think firms complain that France's labour code is too restrictive?**
- 3 **Why might firms want greater flexibility in hiring and firing workers in response to economic conditions?**
- 4 **Do you think that protecting workers, who often have limited market power, is more important than allowing firms to operate flexibly? Explain.**
- 5 **One of the key points of agreement that French trade unions were looking for was for the government to increase the tax on short-term contracts for workers. Why do you think they argued for an increase in this tax?**

SUMMARY

- The demand for labour is a derived demand that comes from firms that use the factors to produce goods and services. Competitive, profit-maximizing firms hire each factor up to the point at which the value of the marginal product of the factor equals its price.
- The supply of labour arises from individuals' trade-off between work and leisure. An upwards sloping labour supply curve means that people respond to an increase in the wage by enjoying less leisure and working more hours.
- The price paid to each factor adjusts to balance the supply and demand for that factor. Because factor demand reflects the value of the marginal product of that factor, in equilibrium each factor is compensated according to its marginal contribution to the production of goods and services.
- Because factors of production are used together, the marginal product of any one factor depends on the quantities of all factors that are available. As a result, a change in the supply of one factor alters the equilibrium earnings of all the factors.
- Workers earn different wages for many reasons. To some extent, wage differentials compensate workers for job attributes. Other things equal, workers in hard, unpleasant jobs get paid more than workers in easy, pleasant jobs.
- Workers with more human capital get paid more than workers with less human capital. The return to accumulating human capital is high and has increased over the past two decades.
- Although years of education, experience and job characteristics affect earnings as theory predicts, there is much variation in earnings that cannot be explained by things that economists can measure. The unexplained variation in earnings is largely attributable to natural ability, effort and chance.
- Some economists have suggested that more educated workers earn higher wages not because education raises productivity but because workers with high natural ability use education as a way to signal their high ability to employers. If this signalling theory is correct, then increasing the educational attainment of all workers would not raise the overall level of wages.
- Wages are sometimes pushed above the level that brings supply and demand into balance. Three reasons for above-equilibrium wages are minimum wage laws, unions and efficiency wages.
- Some differences in earnings are attributable to discrimination on the basis of race, sex or other factors. Measuring the amount of discrimination is difficult, however, because one must correct for differences in human capital and job characteristics.
- Competitive markets tend to limit the impact of discrimination on wages. If the wages of a group of workers are lower than those of another group for reasons not related to marginal productivity, then non-discriminatory firms will be more profitable than discriminatory firms. Profit maximizing behaviour, therefore, can reduce discriminatory wage differentials. Discrimination persists in competitive markets, however, if customers are willing to pay more to discriminatory firms or if the government passes laws requiring firms to discriminate.

QUESTIONS FOR REVIEW

- 1 Explain how a firm's production function is related to its marginal product of labour, how a firm's marginal product of labour is related to the value of its marginal product and how a firm's value of marginal product is related to its demand for labour.
- 2 Give two examples of events that could shift the demand for labour.
- 3 Give two examples of events that could shift the supply of labour.
- 4 Explain how the wage can adjust to balance the supply and demand for labour while simultaneously equalling the value of the marginal product of labour.
- 5 If the population of Norway suddenly grew because of a large immigration, what would you expect to happen to wages? What would happen to the rents earned by the owners of land and capital?
- 6 Why do coal miners get paid more than other workers with similar amounts of education?
- 7 How might education raise a worker's wage without raising the worker's productivity?
- 8 Give three reasons why a worker's wage might be above the level that balances supply and demand.
- 9 What difficulties arise in deciding whether a group of workers has a lower wage because of discrimination?
- 10 Give an example of how discrimination might persist in a competitive market.

PROBLEMS AND APPLICATIONS

- 1 Suppose that the government proposes a new law aimed at reducing health care costs: all citizens are to be required to eat one apple daily.
 - a. How would this apple-a-day law affect the demand and equilibrium price of apples?
 - b. How would the law affect the marginal product and the value of the marginal product of apple pickers?
 - c. How would the law affect the demand and equilibrium wage for apple pickers?
- 2 Show the effect of each of the following events on the market for labour in the tablet manufacturing industry.
 - a. The government buys tablets for all university students.
 - b. More university students graduate in engineering and computer science.
 - c. Computer firms build new manufacturing factories.
- 3 Your enterprising uncle opens a sandwich shop that employs seven people. The employees are paid €6 per hour and a sandwich sells for €13. If your uncle is maximizing his profit, what is the value of the marginal product of the last worker he hired? What is that worker's marginal product?
- 4 Imagine a firm that hires two types of workers – some with computer skills and some without. If technology advances so that computers become more useful to the firm, what happens to the marginal product of the two types? What happens to equilibrium wages? Explain, using appropriate diagrams.
- 5 Suppose a harsh winter in Normandy destroys part of the French apple crop.
 - a. Explain what happens to the price of apples and the marginal product of apple pickers as a result of the freeze. Can you say what happens to the demand for apple pickers? Why or why not?
 - b. Suppose the price of apples doubles and the marginal product falls by 30 per cent. What happens to the equilibrium wage of apple pickers?
 - c. Suppose the price of apples rises by 30 per cent and the marginal product falls by 50 per cent. What happens to the equilibrium wage of apple pickers?
- 6 Suppose that labour is the only input used by a perfectly competitive firm that can hire workers for €150 per day. The firm's production function is as follows:

| Days of labour | Units of output |
|----------------|-----------------|
| 0 | 0 |
| 1 | 7 |
| 2 | 13 |
| 3 | 19 |
| 4 | 25 |
| 5 | 28 |
| 6 | 29 |

Each unit of output sells for €110. Plot the firm's demand for labour. How many days of labour should the firm hire? Show this point on your graph.
- 7 This chapter has assumed that labour is supplied by individual workers acting competitively. In some markets, however, the supply of labour is determined by a union of workers.
 - a. Explain why the situation faced by a labour union may resemble the situation faced by a monopoly firm.
 - b. The goal of a monopoly firm is to maximize profits. Is there an analogous goal for labour unions?

- c. Now extend the analogy between monopoly firms and unions. How do you suppose that the wage set by a union compares to the wage in a competitive market? How do you suppose employment differs in the two cases?
 - d. What other goals might unions have that make unions different from monopoly firms?
- 8** University students sometimes work as summer interns for private firms or the government. Many of these positions pay little or nothing.
- a. What is the opportunity cost of taking such a job?
 - b. Explain why students are willing to take these jobs.
 - c. If you were to compare the earnings later in life of workers who had worked as interns and those who had taken summer jobs that paid more, what would you expect to find?
- 9** A minimum wage law distorts the market for low-wage labour. To reduce this distortion, some economists advocate a two-tiered minimum wage system, with a regular minimum wage for adult workers and a lower, 'sub-minimum' wage for teenage workers. Give two reasons why a single minimum wage might distort the labour market for teenage workers more than it would the market for adult workers.
- 10** This chapter considers the economics of discrimination by employers, customers and governments. Now consider discrimination by workers. Suppose that some brunette workers did not like working with blonde workers. Do you think this worker discrimination could explain lower wages for blonde workers? If such a wage differential existed, what would a profit-maximizing entrepreneur do? If there were many such entrepreneurs, what would happen over time?

PART 8

INEQUALITY

18

INCOME INEQUALITY AND POVERTY

The only difference between the rich and other people,' Mary Colum once said to Ernest Hemingway,' is that the rich have more money.' This may be so but the statement leaves many questions unanswered. The gap between rich and poor is a fascinating and important topic of study – for the very rich, the comfortably rich, for the struggling poor, and indeed for all members of a modern economy. If Hemingway had studied economics, he might have used this reply: 'Yes, but why are they so rich while others are not so rich?'

From the last chapter you should have some understanding about why different people have different incomes. A person's earnings depend on the supply and demand for that person's labour, which in turn depend on natural ability, human capital, compensating differentials, discrimination and so on. The factors that determine wages are also largely responsible for determining how an economy's total income is distributed among the various members of society. In other words, they determine who is rich and who is poor.

In this chapter we discuss the distribution of income – a topic that raises some fundamental questions about the role of economic policy. One of the *Ten Principles of Economics* is that governments can sometimes improve market outcomes. This possibility is particularly important when considering the distribution of income. The invisible hand of the marketplace acts to allocate resources efficiently, but it does not necessarily ensure that resources are allocated fairly. As a result, many economists – although not all – believe that the government should redistribute income to achieve greater equality. In doing so, however, the government runs into another of the *Ten Principles of Economics*: people face trade-offs. When the government enacts policies to make the distribution of income more equitable, it distorts incentives, alters behaviour and makes the allocation of resources less efficient.

Our discussion of the distribution of income proceeds in three steps. First, we assess how much inequality there is in our society. Second, we consider some different views about what role the government should play in altering the distribution of income. Third, we discuss various public policies aimed at helping society's poorest members.

THE MEASUREMENT OF INEQUALITY

We begin our study of the distribution of income by addressing four questions of measurement:

- How much inequality is there in our society?
- How many people live in poverty?
- What problems arise in measuring the amount of inequality?
- How often do people move among income classes?

These measurement questions are the natural starting point from which to discuss public policies aimed at changing the distribution of income.

Income Inequality

There are a number of different ways to measure income inequality. Imagine that you lined up all the families in the economy according to their annual income. Then you divided the families into groups: the bottom 10 per cent, the next 10 per cent and so on until we got to the top 10 per cent. This would be dividing the population by deciles. If you divided the families into groups, the bottom 20 per cent, the next 20 per cent and so on, this would be dividing the population by quintile. We could then look at how incomes fall in these different groupings to get some indication of the income distribution. What we are interested in is how evenly incomes are distributed in an economy. Are the differences in incomes between the lower decile or quintile groups quite similar to those of the top decile or quintile groups or are the figures skewed to one end of the scale, i.e. do the richest 10 per cent or 20 per cent of the population account for a far higher proportion of income than the lower groups? A characteristic of some economies is that a relatively small number of people account for a large proportion of incomes in the economy; in other words, there are a relatively small number of very wealthy people in the economy but a large number of people who are on much lower incomes.

According to data from the Institute for Fiscal Studies, the distribution of income in the UK is highly skewed. The mean income in the UK in 2010–11 was calculated at £511 per week (€599) and the median income (the middle income) was £419 (€491). 65 per cent of individuals in the UK live in households with incomes below £511 per week and only 1.3 million individuals out of a population of over 60 million have household incomes above £1,500 per week (€1,760).

Across Europe there are similar income inequalities. In Greece, for example, the top 20 per cent of the population earn 6.0 per cent more than the bottom 20 per cent, in Bulgaria and Spain it is higher at 6.5 and 6.8 per cent respectively whilst in Iceland and Norway the inequalities are not as large with the top 20 per cent earning 3.3 per cent more than the bottom 20 per cent. Across the EU 27 countries as a whole the ratio of total income received by the top quintile with the highest income to that received by the bottom quintile is 5.1 per cent. (Source of data, Eurostat. This data was collected before Croatia joined the European Union).

There are two common measures of income inequality which we will now look at, the Lorenz curve and the Gini coefficient.

The Lorenz Curve

We noted above that the population (or households) can be grouped in different ways. The **Lorenz curve** shows the relationship between the cumulative percentage of households and the cumulative percentage of income.

Lorenz curve the relationship between the cumulative percentage of households and the cumulative percentage of income

Figure 18.1 shows the relationship in graphical form. If income is distributed evenly, then each proportion of households account for the same percentage of income. The resulting line connecting

all these points is a 45 degree line of perfect equality. If total income in the country was €100 million then the bottom 10 per cent would account for €10 million, the next 10 per cent, €10 million and so on. However, we know that such equality is highly unlikely so the Lorenz curve reflects the degree of inequality in a country.

Assume that the share of income in a country is that given in Table 18.1 measured in quintiles. The bottom 20 per cent (quintile) has a share of total income of 5 per cent, the second quintile, 10 per cent, the third, 25 per cent, the fourth 30 per cent and the top quintile 30 per cent. The share of income must add up to 100 per cent and when we express the share as the cumulative share of income we get the figures in the third column.

TABLE 18.1 The Lorenz Curve

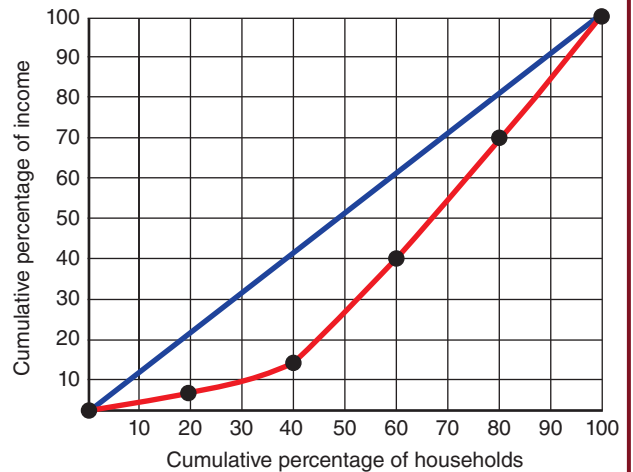
| Quintile | Percentage share of income | Cumulative share of income |
|--------------------|----------------------------|----------------------------|
| Bottom 20 per cent | 5 | 5 |
| Second 20 per cent | 10 | 15 |
| Third 20 per cent | 25 | 40 |
| Fourth 20 per cent | 30 | 70 |
| Top 20 per cent | 30 | 100 |

We then plot the data from the third column onto our graph to get the Lorenz curve shown in Figure 18.1.

FIGURE 18.1

The Lorenz Curve

If each quintile accounted for the same percentage of cumulative income the line of perfect equality would be a 45 degree line. The Lorenz curve shows the degree of inequality in a country – the more bowed the curve the greater the degree of inequality.



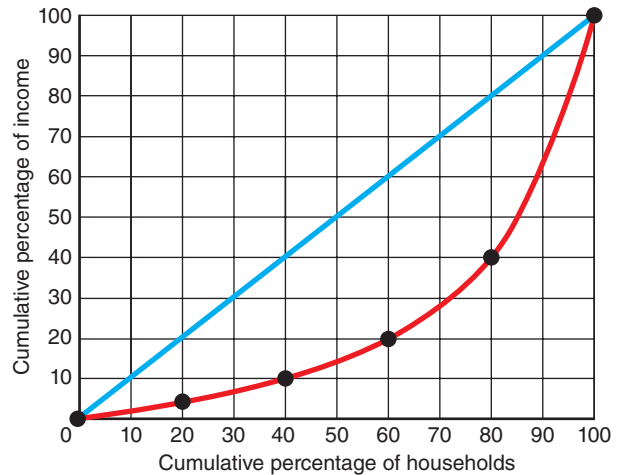
Compare the Lorenz curve in Figure 18.1 with that in Figure 18.2. The cumulative share in income in Figure 18.2 is taken from Table 18.2.

TABLE 18.2 Lorenz Curve Showing Greater Inequality of Income

| Quintile | Percentage share of income | Cumulative share of income |
|--------------------|----------------------------|----------------------------|
| Bottom 20 per cent | 5 | 5 |
| Second 20 per cent | 5 | 10 |
| Third 20 per cent | 10 | 20 |
| Fourth 20 per cent | 20 | 40 |
| Top 20 per cent | 60 | 100 |

FIGURE 18.2**Lorenz Curve Showing Greater Inequality of Income**

The Lorenz curve shown in the figure in comparison to the one in Figure 18.1 is more bowed reflecting the greater degree of income inequality in this country. In this example, income is heavily concentrated in the hands of the top 20 per cent of households.



The Lorenz curve in Figure 18.2 is more bowed than that in Figure 18.1. The reason is that the degree of income inequality in this country is greater than that in the country depicted in Figure 18.1. In this country, the bottom 40 per cent of households only account for 10 per cent of income whereas the top 20 per cent accounts for 60 per cent of total income. The more bowed the Lorenz curve the greater the degree of income inequality.

The Gini Coefficient

We have seen that the more bowed the Lorenz curve the greater the degree of income inequality. The Gini coefficient was developed by an Italian statistician, Corrado Gini (1884–1965) in 1912. The **Gini coefficient** measures the ratio of the area between the 45 degree line of perfect income equality (a benchmark of absolute equality) and the Lorenz curve to the entire area under the 45 degree line of perfect income equality.

$$\text{Gini coefficient} = \frac{\text{Area between the line of perfect income equality and Lorenz curve}}{\text{Area under the line of perfect income equality}}$$

Comparing the Gini coefficient between different countries allows us to observe different income distributions. The Gini coefficient tells us nothing about **how** income is distributed between different countries, merely that one country has a more unequal distribution than another.

Gini coefficient a measure of the degree of inequality of income in a country

The Gini coefficient is a number between 0 and 1. A Gini coefficient of 0 means that income equality is perfect, in other words there is no difference between the line of perfect income equality and the actual Lorenz curve. The numerator in the ratio is, therefore, zero and the denominator 1. At the other extreme if all income was in the hands of just one household then the area between the line of perfect income equality and the actual Lorenz curve would be equal to 1 and the denominator would also be 1. It follows, therefore, that the higher the Gini coefficient the greater the degree of income inequality.

The principle of the Gini coefficient is shown in Figure 18.3.

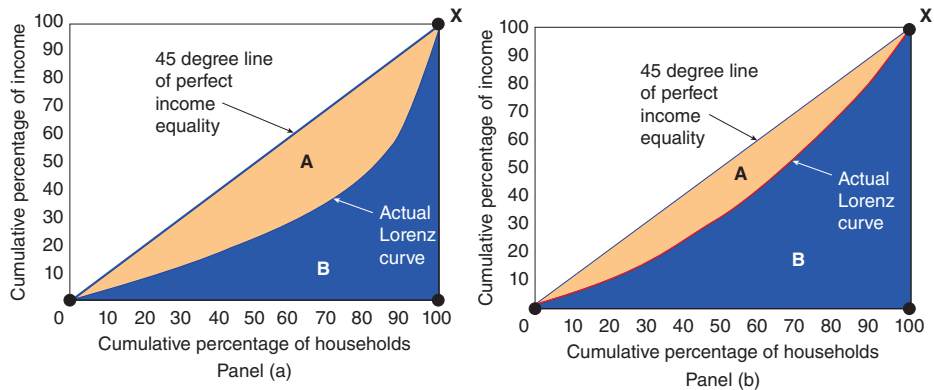
In panel (a) the area between the line of perfect income inequality and the actual Lorenz curve is the orange shaded area A, and the total area under the 45 degree line of perfect income inequality is the triangle 0,X,100 (area A plus the blue shaded area, B). Calculating the area A and dividing that by the total

area A + B, gives the Gini coefficient. Calculating the area between the 45 degree line and the actual Lorenz curve is done through the use of integral calculus and is described in detail in the maths supplement to this book. Panel (b) shows a situation where the area A is much smaller than that shown for the country in panel (a) and in this case the income inequality would be much less reflected by a lower Gini coefficient.

FIGURE 18.3

The Gini Coefficient

The Gini coefficient is found by dividing the area A by the total area under the 45 degree line of perfect income equality (the area A + B).

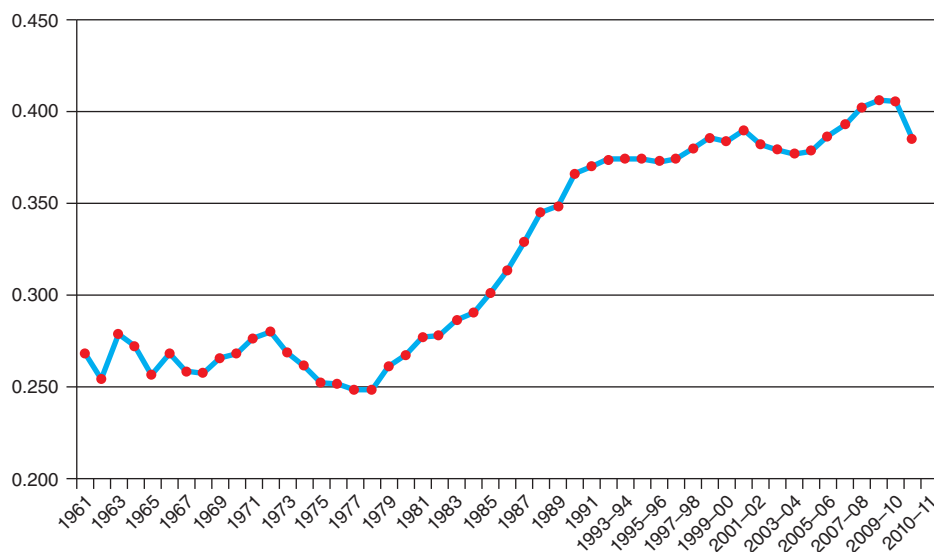


Gini Coefficients in the UK and Europe Figure 18.4 shows the change in the Gini coefficient for Great Britain between 1961 and 2010–2011. It can be seen that the degree of income inequality has risen quite markedly over the last 50 years from around 0.25 in the 1960s to over 0.4 in the later part of the first decade of the 21st century, falling back to around 0.38 in 2010–2011 (Source: Institute for Fiscal Studies).

FIGURE 18.4

Gini Coefficient in Great Britain Between 1961 and 2010–2011

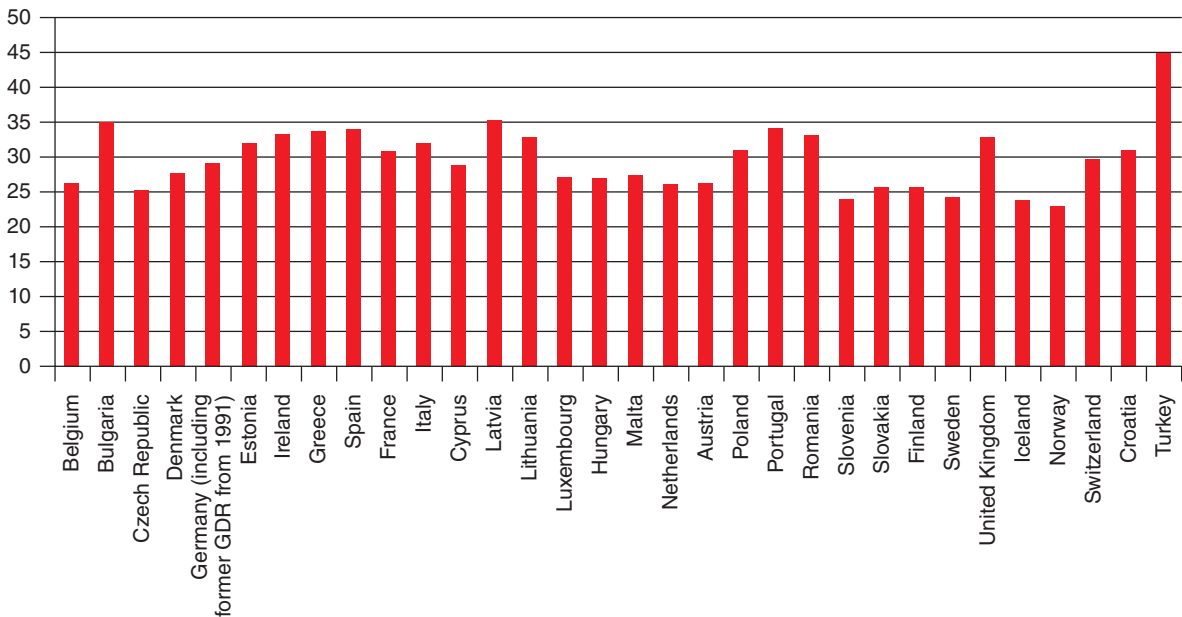
Over the last 50 years the Gini coefficient in Great Britain has increased suggesting that the inequality of income has increased in the country.



The Gini coefficient for the 27 countries of the EU, as a whole, expressed on a scale from 0–100 was 30.7 in 2011, a slight increase from 30.6 in 2005. Norway is one of the countries with the lowest Gini coefficient at 22.9 in 2011, whereas Turkey is one of the highest at 44.8 (although this is a figure from 2006).

FIGURE 18.5**Gini Coefficient (Measured on a Scale from 0–100) for European Countries, 2011**

The Gini coefficients as measured on a scale from 0–100 for European countries, 2011. Note: the figure for Turkey is from 2006 and Ireland from 2010.



Problems in Measuring Inequality

Although data on the income distribution give us some idea about the degree of inequality in our society, interpreting these data is not as straightforward as it might first appear. The data are based on households' annual incomes. What people care about, however, is not their incomes but their ability to maintain a good standard of living. For various reasons, data on the income distribution give an incomplete picture of inequality in living standards. We examine these reasons below.

The Economic Life Cycle Incomes vary predictably over people's lives. A young worker, especially one still engaged in full-time study, has a low income. Income rises as the worker gains maturity and experience, peaks at around age 50, and then falls sharply when the worker retires at around age 65. This regular pattern of income variation is called the **life cycle**.

life cycle the regular pattern of income variation over a person's life

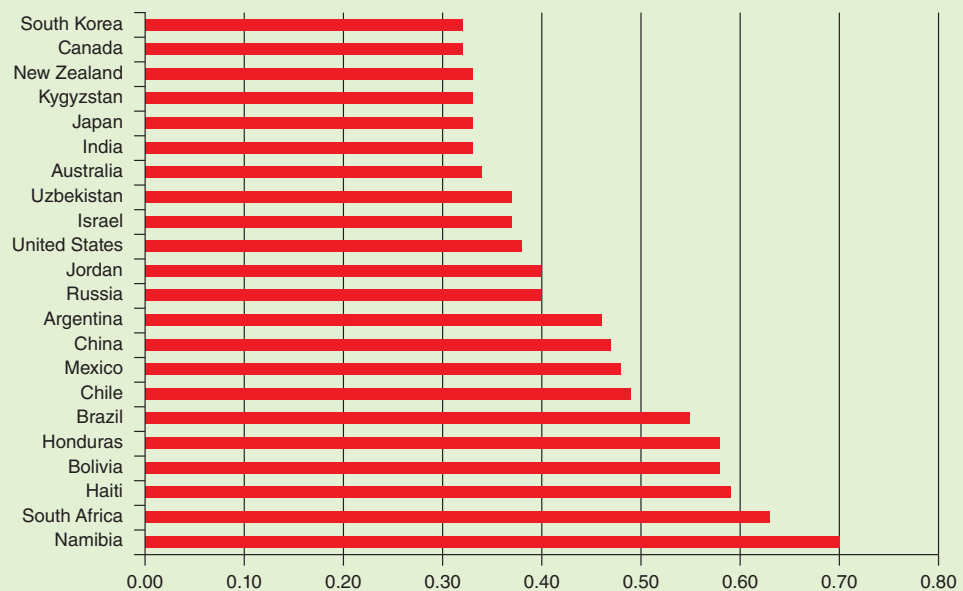
CASE STUDY Gini coefficients Around the World

Having looked at the Gini coefficients in Europe, how do these compare with those across the rest of the world? Figure 18.6 shows some selected Gini coefficients from around the world (data sourced from the OECD and the *CIA Factbook*). One interesting thing to note about the data is the variability in how current the data is. Collecting data on income inequality is not easy because not every country has advanced statistical services that can collect and publish such data. In Figure 18.6, some of the data is from 2001 (Haiti), 2007 (Honduras) and 2009 (Bolivia and Argentina). Despite the differences in dates from which the data are sourced, the changes in income inequality tend to be relatively slow so even data from 10 years ago can still be useful in giving a guide to comparing income inequalities across different countries.

FIGURE 18.6

Gini Coefficients from Around the World

The figure shows Gini coefficients from selected countries around the world arranged in descending order.



Both Namibia and South Africa have relatively high Gini coefficients reflecting in South Africa's case the legacy of Apartheid where many of the country's poorest citizens are still struggling to close the gap on the rich. There are also relatively high Gini coefficients in South American countries. Brazil is termed an emerging economy, one of the so called BRIC



nations (Brazil, Russia, India and China) but whilst economic growth in Brazil has been relatively strong, the degree of inequality in the country is relatively high. If we were to draw a comparison with the 28 countries of the EU, it could be concluded that across the EU the degree of income inequality is relatively low compared to that across the rest of the world.

Countries with relatively high Gini coefficients have a vast gulf between the rich and the poor.

Because people can borrow and save to smooth out life cycle changes in income, their standard of living in any year depends more on lifetime income than on that year's income. The young often borrow, perhaps to go to university or to buy a house, and then repay these loans later when their incomes rise. People have their highest saving rates when they are middle-aged. Because people can save in anticipation of retirement, the large declines in incomes at retirement need not lead to similar declines in standard of living.

This normal life-cycle pattern causes inequality in the distribution of annual income, but it does not represent true inequality in living standards. To gauge the inequality of living standards in our society, the distribution of lifetime incomes is more relevant than the distribution of annual incomes. Unfortunately, data on lifetime incomes are not readily available. When looking at any data on inequality, however, it is important to keep the life cycle in mind. Because a person's lifetime income smoothes out the highs and lows of the life cycle, lifetime incomes are surely more equally distributed across the population than are annual incomes.

Transitory versus Permanent Income Incomes vary over people's lives not only because of predictable life-cycle variation but also because of random and transitory forces. One year a frost kills off the Normandy apple crop and Normandy apple growers see their incomes fall temporarily. At the same time, the Normandy frost drives up the price of apples and English apple growers see their incomes temporarily rise. The next year the reverse might happen.

Just as people can borrow and lend to smooth out life-cycle variation in income, they can also borrow and lend to smooth out transitory variation in income. When English apple growers experience a good year, they would be foolish to spend all of their additional income. Instead, they save some of it, knowing that their good fortune is unlikely to persist. Similarly, the Normandy apple growers respond to their temporarily low incomes by drawing down their savings or by borrowing. To the extent that a family saves and borrows to buffer itself from transitory changes in income, these changes do not affect its standard of living. A family's ability to buy goods and services depends largely on its **permanent income**, which is its normal, or average, income.

permanent income a person's normal income

To gauge inequality of living standards, the distribution of permanent income is more relevant than the distribution of annual income. Although permanent income is hard to measure, it is an important concept. Because permanent income excludes transitory changes in income, permanent income is more equally distributed than is current income.

Economic Mobility

People sometimes speak of 'the rich' and 'the poor' as if these groups consisted of the same families year after year. In fact, this is not at all the case. Economic mobility, the movement of people among income classes, is quite substantial in European and North American economies. Movements up the income ladder can be due to good luck or hard work, and movements down the ladder can be due to bad luck or laziness. Some of this mobility reflects transitory variation in income, while some reflects more persistent changes in income.

Because economic mobility is so great, poverty is a long-term problem for relatively few families. Because it is likely that the temporarily poor and the persistently poor face different problems, policies that aim to combat poverty need to distinguish between these groups.

Another way to gauge economic mobility is the persistence of economic success from generation to generation. Economists who have studied this topic find substantial mobility. If a father earns

20 per cent above his generation's average income, his son will most likely earn 8 per cent above his generation's average income. There is almost no correlation between the income of a grandfather and the income of a grandson.

The Poverty Rate

A commonly used gauge of the distribution of income is the poverty rate. The **poverty rate** is the percentage of the population whose family income falls below an absolute level called the **poverty line**. Poverty is a relative concept – one person's poverty is another person's wealthy. A millionaire is wealthy compared to someone earning €50,000 a year but is poor compared to a billionaire! For this reason economists distinguish between absolute and relative poverty. **Absolute poverty** is when individuals do not have access to the basic requirements of life – food, shelter and clothing. **Relative poverty** occurs when individuals are excluded from being able to take part in what are considered the normal, acceptable standards of living in a society.

poverty rate the percentage of the population whose family income falls below an absolute level called the poverty line

poverty line an absolute level of income set by the government below which a family is deemed to be in poverty. In the UK and Europe this is measured by earnings less than 60 per cent of median income

absolute poverty a level of poverty where an individual does not have access to the basics of life – food, clothing and shelter

relative poverty a situation where an individual is not able to access what would be considered acceptable standards of living in society

To get some idea of levels of poverty across different countries, therefore, similar measures have to be used. In Europe the measure for defining the poverty line is set at 60 per cent of the median income. If measuring across Europe, this income has to be equalized and is called the equalized household income and takes in to account differences in the cost of living across the EU. If the median income is €20,000 then any family earning less than €12,000 a year would be classed as living in poverty. In the UK there are over 12 million families classed as living in poverty. The median income in different European countries varies. The EU statistics service, Eurostat, publishes data showing at-risk-of-poverty thresholds for single people set at 60 per cent of the national median income. Figure 18.7 shows these thresholds. It is clear that these vary considerably across Europe with Turkey's being very low (based on 2006 figures) at €1,424, whereas in Malta the threshold is €6,517 and in Switzerland it is €20,362.

The implications of these figures is startling. Some 24 per cent of the population of the EU is at risk of poverty (equivalent to almost 120 million people). The percentage of the population at risk of poverty across Europe is shown in Figure 18.8.

The highest risk countries tend to be those that were part of the former Soviet Union which have been undergoing considerable economic restructuring from command economies to market economies. In Bulgaria, for example, just over 49 per cent of the population is at risk of poverty, in Latvia and Romania it is just over 40 per cent and in Lithuania and Hungary, 33.4 per cent and 31 per cent respectively.

Looking at the poverty rate in addition to other data on inequality is important. We might observe that average incomes have risen over time but not everyone is able to share in the increased prosperity that has occurred. Understanding something of why some people get left behind is crucial to developing policies to support people to help them gain a better standard of living.

FIGURE 18.7

At Risk of Poverty Thresholds for European Countries 2011

The variation in the at-risk-of-poverty thresholds across Europe varies. The figure for Turkey is for 2006 and for Ireland is for 2010.

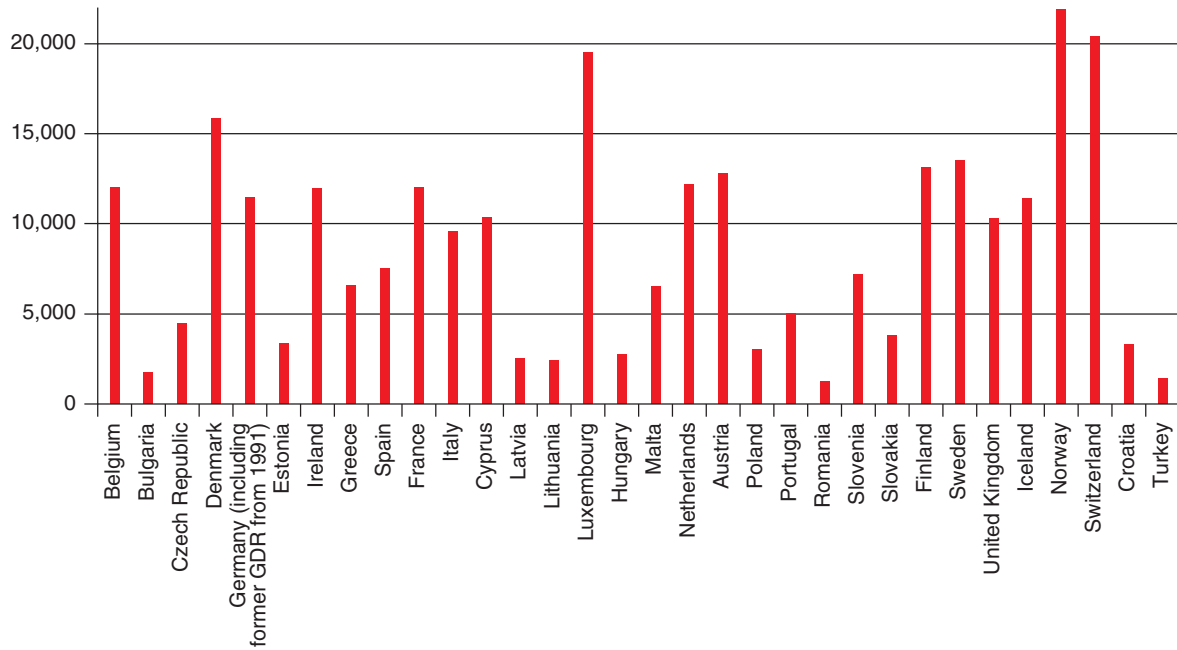
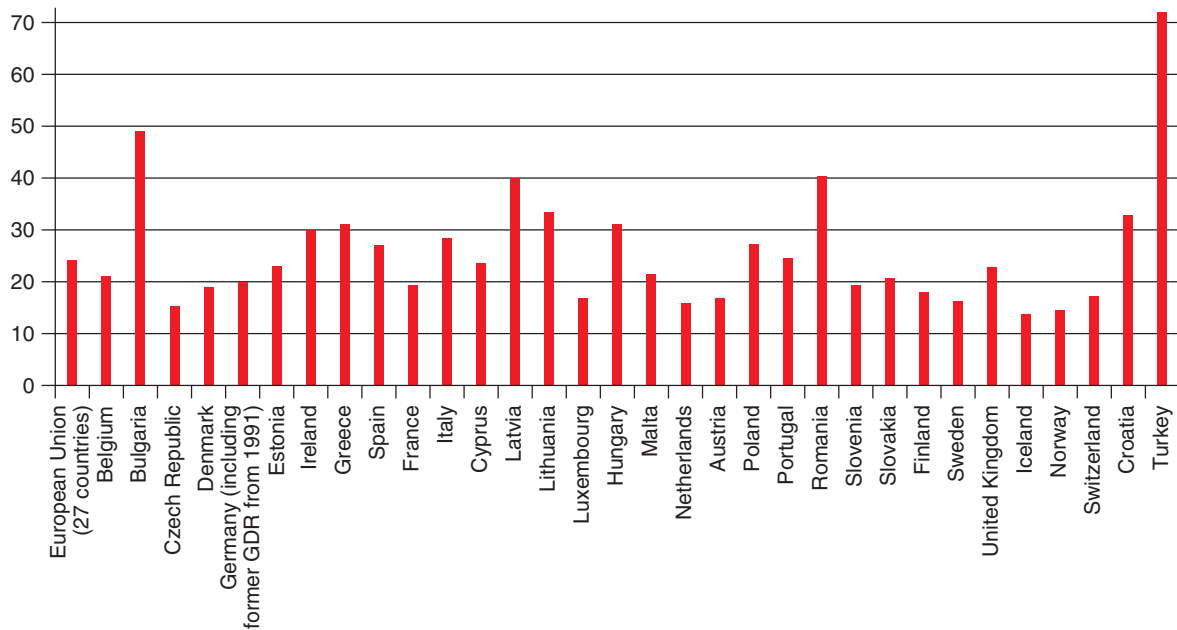


FIGURE 18.8

People at Risk of Poverty and Social Exclusion in Europe in 2011

The percentage of people at risk of poverty across Europe varies. The figure for Turkey is from 2006 and for Ireland, from 2010. Turkey has the highest percentage at risk at over 70 per cent. The former Soviet states also have a relatively high risk of poverty but across the EU 27 states as a whole almost a quarter of the population are at risk of poverty.



THE POLITICAL PHILOSOPHY OF REDISTRIBUTING INCOME

We have just seen how the economy's income is distributed and have considered some of the problems in interpreting measured inequality. This discussion was *positive* in the sense that it merely described the world as it is. We now turn to the *normative* question facing policymakers: what *should* the government do about economic inequality?

This question is not just about economics. Economic analysis alone cannot tell us whether policymakers *should* try to make our society more egalitarian. Our views on this question are, to a large extent, a matter of political philosophy. Yet because the government's role in redistributing income is central to so many debates over economic policy, here we digress from economic science to consider a bit of political philosophy.

Utilitarianism

A prominent school of thought in political philosophy is **utilitarianism**. The founders of utilitarianism were the British philosophers Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873). To a large extent, the goal of utilitarians is to apply the logic of individual decision making to questions concerning morality and public policy.

utilitarianism the political philosophy according to which the government should choose policies to maximize the total utility of everyone in society

The starting point of utilitarianism is the notion of utility, which, you might recall is the level of happiness or satisfaction that a person receives from consumption or his or her circumstances. Utility is a measure of well-being and, according to utilitarians, is the ultimate objective of all public and private actions. The proper goal of the government, they claim, is to maximize the sum of utility of everyone in society.

The utilitarian case for redistributing income is based on the assumption of *diminishing marginal utility*. It seems reasonable that an extra euro of income to a poor person provides that person with more additional utility than does an extra euro to a rich person. In other words, as a person's income rises, the extra well-being derived from an additional euro of income falls. This plausible assumption, together with the utilitarian goal of maximizing total utility, implies that the government should try to achieve a more equal distribution of income.

The argument is simple. Imagine that Dieter and Ernst are the same, except that Dieter earns €80,000 and Ernst earns €20,000. In this case, taking a euro from Dieter to pay Ernst will reduce Dieter's utility and raise Ernst's utility. But, because of diminishing marginal utility, Dieter's utility falls by less than Ernst's utility rises. Thus, this redistribution of income raises total utility, which is the utilitarian's objective.

At first, this utilitarian argument might seem to imply that the government should continue to redistribute income until everyone in society has exactly the same income. Indeed, that would be the case if the total amount of income – €100,000 in our example – were fixed. But, in fact, it is not. Utilitarians reject complete equalization of incomes because they accept one of the *Ten Principles of Economics* presented: people respond to incentives.

To take from Dieter to pay Ernst, the government must pursue policies that redistribute income, such as the income tax and welfare systems that operate in all industrialized countries. Under these policies, people with high incomes pay high taxes, and people with low incomes receive income transfers. We know, however, that taxes distort incentives and cause deadweight losses. If the government takes away additional income a person might earn through higher income taxes or reduced transfers, both Dieter

and Ernst have less incentive to work hard. As they work less, society's income falls, and so does total utility. The utilitarian government has to balance the gains from greater equality against the losses from distorted incentives. To maximize total utility, therefore, the government stops short of making society fully egalitarian.

A famous parable sheds light on the utilitarian's logic. Imagine that Dieter and Ernst are thirsty travellers trapped at different places in the desert. Dieter's oasis has much water; Ernst's has little. If the government could transfer water from one oasis to the other without cost, it would maximize total utility from water by equalizing the amount in the two places. But suppose that the government has only a leaky bucket. As it tries to move water from one place to the other, some of the water is lost in transit. In this case, a utilitarian government might still try to move some water from Dieter to Ernst, depending on how thirsty Ernst is and how leaky the bucket is. But, with only a leaky bucket at its disposal, a utilitarian government will not try to reach complete equality.

Liberalism

A second way of thinking about inequality might be called **liberalism**. Philosopher John Rawls develops this view in his book *A Theory of Justice*. This book was first published in 1971, and it quickly became a classic in political philosophy.

liberalism the political philosophy according to which the government should choose policies deemed to be just, as evaluated by an impartial observer behind a 'veil of ignorance'

Rawls begins with the premise that a society's institutions, laws and policies should be just. He then takes up the natural question: how can we, the members of society, ever agree on what justice means? It might seem that every person's point of view is inevitably based on his or her particular circumstances – whether he or she is talented or less talented, diligent or lazy, educated or less educated, born to a wealthy family or a poor one. Could we ever *objectively* determine what a just society would be?

To answer this question, Rawls proposes the following thought experiment. Imagine that before any of us is born, we all get together for a meeting to design the rules that govern society. At this point, we are all ignorant about the station in life each of us will end up filling. In Rawls's words, we are sitting in an 'original position' behind a 'veil of ignorance'. In this original position, Rawls argues, we can choose a just set of rules for society because we must consider how those rules will affect every person. As Rawls puts it, 'Since all are similarly situated and no one is able to design principles to favour his particular conditions, the principles of justice are the result of fair agreement or bargain.' Designing public policies and institutions in this way allows us to be objective about what policies are just.

Rawls then considers what public policy designed behind this veil of ignorance would try to achieve. In particular, he considers what income distribution a person would consider fair if that person did not know whether he or she would end up at the top, bottom or middle of the distribution. Rawls argues that a person in the original position would be especially concerned about the possibility of being at the *bottom* of the income distribution. In designing public policies, therefore, we should aim to raise the welfare of the worst-off person in society. That is, rather than maximizing the sum of everyone's utility, as a utilitarian would do, Rawls would maximize the minimum utility. Rawls's rule is called the **maximin criterion**.

maximin criterion the claim that the government should aim to maximize the well-being of the worst-off person in society

Because the maximin criterion emphasizes the least fortunate person in society, it justifies public policies aimed at equalizing the distribution of income. By transferring income from the rich to the poor, society raises the well-being of the least fortunate. The maximin criterion would not, however, lead to a completely egalitarian society. If the government promised to equalize incomes completely, people would have no incentive to work hard, society's total income would fall substantially and the least fortunate person would be worse off. Thus, the maximin criterion still allows disparities in income, because such disparities can improve incentives and thereby raise society's ability to help the poor. Nonetheless, because Rawls's philosophy puts weight on only the least fortunate members of society, it calls for more income redistribution than does utilitarianism.

Rawls's views are controversial, but the thought experiment he proposes has much appeal. In particular, this thought experiment allows us to consider the redistribution of income as a form of *social insurance*. That is, from the perspective of the original position behind the veil of ignorance, income redistribution is like an insurance policy. Homeowners buy fire insurance to protect themselves from the risk of their housing burning down. Similarly, when we as a society choose policies that tax the rich to supplement the incomes of the poor, we are all insuring ourselves against the possibility that we might have been a member of a poor family. Because people dislike risk, we should be happy to have been born into a society that provides us this insurance.

It is not at all clear, however, that rational people behind the veil of ignorance would truly be so averse to risk as to follow the maximin criterion. Indeed, because a person in the original position might end up anywhere in the distribution of outcomes, he or she might treat all possible outcomes equally when designing public policies. In this case, the best policy behind the veil of ignorance would be to maximize the average utility of members of society, and the resulting notion of justice would be more utilitarian than Rawlsian.

Libertarianism

A third view of inequality is called **libertarianism**. The two views we have considered so far – utilitarianism and liberalism – both view the total income of society as a shared resource that a social planner can freely redistribute to achieve some social goal. By contrast, libertarians argue that society itself earns no income – only individual members of society earn income. According to libertarians, the government should not take from some individuals and give to others in order to achieve any particular distribution of income.

libertarianism the political philosophy according to which the government should punish crimes and enforce voluntary agreements but not redistribute income

For instance, philosopher Robert Nozick writes the following in his famous 1974 book *Anarchy, State, and Utopia*:

We are not in the position of children who have been given portions of pie by someone who now makes last minute adjustments to rectify careless cutting. There is no central distribution, no person or group entitled to control all the resources, jointly deciding how they are to be doled out. What each person gets, he gets from others who give to him in exchange for something, or as a gift. In a free society, diverse persons control different resources, and new holdings arise out of the voluntary exchanges and actions of persons.

Whereas utilitarians and liberals try to judge what amount of inequality is desirable in a society, Nozick denies the validity of this very question.

The libertarian alternative to evaluating economic *outcomes* is to evaluate the *process* by which these outcomes arise. When the distribution of income is achieved unfairly – for instance, when one person steals from another – the government has the right and duty to remedy the problem. But, as long as the process determining the distribution of income is just, the resulting distribution is fair, no matter how unequal.

Nozick criticizes Rawls's liberalism by drawing an analogy between the distribution of income in society and the distribution of marks awarded to students taking a course of study. Suppose you were asked to judge the fairness of the marks awarded in the economics course you are now taking. Would you imagine yourself behind a veil of ignorance and choose a marks distribution without knowing the talents and efforts of each student? Or would you ensure that the process of assigning marks to students is fair without regard for whether the resulting distribution is equal or unequal? For the case of course marks at least, the libertarian emphasis on process over outcomes is compelling.

Libertarians conclude that equality of opportunities is more important than equality of incomes. They believe that the government should enforce individual rights to ensure that everyone has the same opportunity to use his or her talents and achieve success. Once these rules of the game are established, the government has no reason to alter the resulting distribution of income.

Libertarian Paternalism

We finally introduce a relatively new concept linked to these philosophies put forward by University of Chicago economists, Richard H. Thaler and Cass R. Sunstein (see Thaler, R.H. and Sunstein C.R. (2009) *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin). Libertarian paternalism recognizes that people should be free to choose but that 'choice architects' (the government in the case of making decisions about rectifying inequality) have a legitimate role in trying to influence people's behaviour in order to make their lives longer, healthier and better – improving their utility. Thaler and Sunstein question whether specific policy moves are the best way of changing behaviour to improve utility and whether 'nudges' could achieve the end result whilst retaining the freedom of people to make choices. Nudges relate to details that might often seem insignificant but when paid attention to, can influence the choices people make to 'nudge' them in the direction of improving their own and society's welfare. Their work covers diverse areas but includes savings for pensions and social security systems, both of which have an impact on inequality in society.

SELF TEST Franz earns more than Paloma. Someone proposes taxing Franz in order to supplement Paloma's income. How would a utilitarian, a liberal and a libertarian evaluate this proposal?

Policies to Reduce Poverty

As we have just seen, political philosophers and economists hold various views about what role the government should take in altering the distribution of income. Political debate among the larger population of voters reflects a similar disagreement. Despite these continuing debates, however, most people believe that, at the very least, the government should try to help those most in need. According to a popular metaphor, the government should provide a 'safety net' to prevent any citizen from falling too far.

Poverty is one of the most difficult problems that policymakers face. Poor families are more likely than the overall population to experience homelessness, drug dependency, domestic violence, health problems, teenage pregnancy, illiteracy, unemployment and low educational attainment. Members of poor families are both more likely to commit crimes and more likely to be victims of crimes. Although it is hard to separate the causes of poverty from the effects, there is no doubt that poverty is associated with various economic and social ills.

Suppose that you were a policymaker in the government, and your goal was to reduce the number of people living in poverty. How would you achieve this goal? Here we consider some of the policy options that you might consider. Although each of these options does help some people escape poverty, none of them is perfect, and deciding upon the best combination is not easy.

Minimum Wage Laws

We have looked at minimum wage laws in Chapter 17. To reiterate, advocates view the minimum wage as a way of helping the working poor without any cost to the government. Critics view it as hurting those it is intended to help. For workers with low levels of skill and experience, a high minimum wage forces the wage above the level that balances supply and demand. It therefore raises the cost of labour to firms and

reduces the quantity of labour that those firms demand. The result is higher unemployment among those groups of workers affected by the minimum wage. Although those workers who remain employed benefit from a higher wage, those who might have been employed at a lower wage are worse off.

The magnitude of these effects, we know, depends on the elasticity of demand. Advocates of a high minimum wage argue that the demand for unskilled labour is relatively inelastic, so that a high minimum wage depresses employment only slightly. Critics of the minimum wage argue that labour demand is more elastic, especially in the long run when firms can adjust employment and production more fully. They also note that many minimum wage workers are teenagers from middle-income families, so that a high minimum wage is imperfectly targeted as a policy for helping the poor.

The effects are also dependent on the degree of substitutability between workers in different industries – the ease with which workers can transfer from one industry to another. Minimum wage laws affect all businesses in all industries in different ways; some industries are not affected greatly by minimum wage laws because they already pay in excess of the minimum wage and so labour market equilibrium is not affected in that particular market. In low-paid industries, such as cleaning, hotel and catering and restaurants, for example, all employers are affected in the same way if they have to increase pay to meet minimum wage laws. As a result, no one employer can necessarily gain any advantage over another by paying workers lower wages and thus having lower costs. Whilst minimum wage laws are a contentious issue, it is also a highly complex one requiring detailed analysis and an understanding that the labour market is not simply an amorphous ‘one’; it consists of many smaller markets each of which has a varying influence on other markets.

Social Security

One way to raise the living standards of the poor is for the government of a country to supplement their incomes. The primary way in which the government does this is through **social security**. This is a broad term that generally encompasses various government benefits. In the UK, for example, Income Support is paid mainly to poor people who are either lone parents and carers, or incapable of work, or else disabled, Tax Credits are payments made to those in work and with families but who receive low incomes and have to care for children, while Jobseeker’s Allowance is paid to unemployed people who are able and willing to work but temporarily cannot find a job.

social security government benefits that supplement the incomes of the needy

A common criticism of the social security system is that it may create bad incentives. For example, since a single mother may lose Income Support if she were married to a man who was in work, the Income Support programme may encourage families to break up or encourage illegitimate births. However, governments may introduce other mechanisms to create good incentives. Tax credits, for example, are designed to encourage family members to take paid work in the belief that a person in work, albeit on a low income, is in a better position than one who does not take a job.

How severe are these potential problems with the benefit system? No one knows for sure. Proponents of the benefit system point out that being a poor, single mother is a difficult existence at best, and they are sceptical that many people would be encouraged to pursue such a life if it were not thrust upon them. Moreover, if it can be proved that a person is incapable of work or is disabled, it seems cruel and ridiculous to argue that this is because of the benefits they are receiving. It is often easy, however, for the popular press to portray examples of those who abuse the system; it is important that, as a budding economist, you ask appropriate questions and try to distinguish between fact and opinion in such cases.

Negative Income Tax

Whenever the government chooses a system to collect taxes, it affects the distribution of income. This is clearly true in the case of a progressive income tax, whereby high-income families pay a larger percentage of their income in taxes than do low-income families. Many economists have advocated supplementing the income of the poor using a **negative income tax**. According to this policy, every family would report its income to the government. High-income families would pay a tax based on their incomes. Low-income families would receive a subsidy. In other words, they would ‘pay’ a ‘negative tax’.

negative income tax a tax system that collects revenue from high-income households and gives transfers to low-income households

For example, suppose the government used the following formula to compute a family's tax liability:

$$\text{Taxes due} = (1/3 \text{ of income}) - \text{€}10,000$$

In this case, a family that earned €60,000 would pay €10,000 in taxes, and a family that earned €90,000 would pay €20,000 in taxes. A family that earned €30,000 would owe nothing and a family that earned €15,000 would 'owe' –€5,000. In other words, the government would send this family a cheque for €5,000.

Under a negative income tax, poor families would receive financial assistance without having to demonstrate need. The only qualification required to receive assistance would be a low income. Depending on one's point of view, this feature can be either an advantage or a disadvantage. On the one hand, a negative income tax does not encourage illegitimate births and the break-up of families, as some critics of welfare systems believe current policy does. On the other hand, a negative income tax would subsidize those who are simply lazy and, in some people's eyes, undeserving of government support.

In-Kind Transfers

Another way to help the poor is to provide them directly with some of the goods and services they need to raise their living standards. For example, charities provide the needy with food, shelter and toys at Christmas. Governments in some countries give poor families vouchers that can be used to buy food or clothing in shops; the shops then redeem the vouchers for money. In the UK, poor people may qualify for free school meals for their children and medical benefits such as free prescriptions, dental treatment and eyesight tests.

Is it better to help the poor with these **in-kind transfers** or with direct cash payments? There is no clear answer.

in-kind transfers transfers to the poor given in the form of goods and services rather than cash

Advocates of in-kind transfers argue that such transfers ensure that the poor get what they need most. Among the poorest members of society, alcohol and drug addiction is more common than it is in society as a whole. By providing the poor with food and shelter rather than cash, society can be more confident that it is not helping to support such addictions.

Advocates of cash payments, on the other hand, argue that in-kind transfers are inefficient and disrespectful. The government does not know, they say, what goods and services the poor need most. Many of the poor are ordinary people down on their luck. Despite their misfortune, they are in the best position to decide how to raise their own living standards. Rather than giving the poor in-kind transfers of goods and services that they may not want, it may be better to give them cash and allow them to buy what they think they need most.

Anti-Poverty Policies and Work Incentives

Many policies aimed at helping the poor can have the unintended effect of discouraging the poor from escaping poverty on their own. To see why, consider the following example. Suppose that a family needs an income of €15,000 a year to maintain a reasonable standard of living. And suppose that, out of concern for the poor, the government promises to guarantee every family that income. Whatever a family earns, the government makes up the difference between that income and €15,000. What effect would you expect this policy to have?

The incentive effects of this policy are obvious: any person who would earn under €15,000 by working has no incentive to find and keep a job. For every euro that the person would earn, the government would reduce the income supplement by one euro. In effect, the government taxes 100 per cent of additional earnings. An effective marginal tax rate of 100 per cent is surely a policy with a large deadweight loss.

The adverse effects of this high effective tax rate can persist over time. A person discouraged from working loses the on-the-job training that a job might offer. In addition, his or her children miss the lessons learned by observing a parent with a full-time job, and this may adversely affect their own ability to find and hold a job.

Although the anti-poverty policy we have been discussing is hypothetical, it is not as unrealistic as it might first appear. In the UK, for example, Income Support and Housing Benefit (a subsidy to a poor family's rent or mortgage interest payments) are benefits aimed at helping the poor and are tied to family income. As a family's income rises, the family becomes ineligible for these benefits. When all benefits being received are taken together, it is common for families to face effective marginal tax rates that are very high. Sometimes the effective marginal tax rates even exceed 100 per cent, so that poor families are worse off when they earn more: they are caught in a 'poverty trap'. By trying to help the poor, the government discourages those families from working. According to critics of anti-poverty policies, these social security benefits alter work attitudes and create a 'culture of poverty'.

It might seem that there is an easy solution to this problem: reduce benefits to poor families more gradually as their incomes rise. For example, if a poor family loses €0.30 of benefits for every extra €1 it earns, then it faces an effective marginal tax rate of 30 per cent. Although this effective tax reduces work effort to some extent, it does not eliminate the incentive to work completely.

The problem with this solution is that it greatly increases the cost of the social security system. The more gradual the phase-out, the more families are eligible for some level of benefits – and the more the social security system costs. Thus, policymakers face a trade-off between burdening the poor with high effective marginal tax rates and burdening taxpayers with a costly anti-poverty programme.

There are various other ways to try to reduce the work disincentive of anti-poverty programmes, such as stopping or reducing benefits to people who have not found a job within a reasonable period of time or who have turned down job offers for no good reason. In the UK, this kind of reasoning underpins the structure of the benefits paid to the unemployed: unemployed people may be eligible to receive a benefit called Jobseeker's Allowance. To receive the allowance, however, the claimant must be capable of starting work immediately and of actively taking steps to find a job such as attending interviews, writing applications, improving their skills or seeking job information. They must also have a current 'jobseeker's agreement' with the Employment Service, which includes such information as hours available for work, the desired job and any steps that the claimant is willing to take to find work (such as moving to a different town). They must be prepared to work up to 40 hours a week and have a reasonable prospect of finding work (i.e. not place too many restrictions on the type of work they are willing to undertake). If a claimant refuses to take up a job offer without good reason, they may be denied further payments of Jobseeker's Allowance.

SELF TEST List three policies aimed at helping the poor, and discuss the pros and cons of each.

CONCLUSION

People have long reflected on the distribution of income in society. Plato, the ancient Greek philosopher, concluded that in an ideal society the income of the richest person would be no more than four times the income of the poorest person. Although the measurement of inequality is difficult, it is clear that our society has much more inequality than Plato recommended.

One of the *Ten Principles of Economics* is that governments can sometimes improve market outcomes. There is little consensus, however, about how this principle should be applied to the distribution of income. Philosophers and policymakers today do not agree on how much income inequality is desirable, or even whether public policy should aim to alter the distribution of income. Much of public debate reflects this disagreement. Whenever taxes are raised, for instance, politicians argue over how much of the tax hike should fall on the rich, people from the middle-income group, and the poor.

Another of the *Ten Principles of Economics* is that people face trade-offs. This principle is important to keep in mind when thinking about economic inequality. Policies that penalize the successful and reward the unsuccessful reduce the incentive to succeed. Thus, policymakers face a trade-off between equality and efficiency. The more equally the pie is divided, the smaller the pie becomes. This is the one lesson concerning the distribution of income about which almost everyone agrees.

IN THE NEWS



The Consequences of Income Inequality

Income inequality can lead to a number of different consequences and is one of the reasons why governments around the world express a desire to reduce income inequalities. In this article, the effects of income inequality are linked with the rate of suicides in young men in Ireland.

Concerns Over Rate of Young Male Suicides

The Young Men and Suicide project found that, over the past decade, the rate of deaths from suicide has been five times higher in males than in females in Ireland.

It is estimated that 165 teenagers and young men took their own lives in Ireland in 2011, while 72 teenagers and young men took their own lives in the North.

Although the rate of male suicide in Ireland is relatively low within the EU overall, the rate among young males is among the highest in the EU.

The jointly funded cross-border report said the recent spike in suicide rates in both Ireland and the North coincides with the economic downturn, increasing levels of unemployment, and alcohol consumption. Hanging was found to be the most frequent method of suicide, particularly among young men. The report also found that deliberate self-harm, traditionally more common in females, was now higher among young males than young females. Noel Richardson, lead author of the report, said there was a public and moral requirement for action on the issue of suicide on the island of Ireland.

'There can be no quick-fix solutions to tackling the very grave statistics on suicide in young men on the island of Ireland,' said Dr Richardson. 'But neither is there any place for inertia or ambivalence. There is both a public health and a moral requirement to act.'

Dr Richardson said there needs to be 'a concerted effort to engage more effectively and in a more sustained way with young men, and to plan services and programmes with young men in mind'. He said the report provided a 'blueprint and a roadmap' for action. The report found that the factors most consistently associated with the rise in young male suicide are income inequality, family relationship difficulties, peer relationship problems, school failure, low self-esteem, and violence. Some of the suicide prevention methods highlighted as being effective in reducing suicide rates include physician education in depression recognition and treatment, and restricting access to lethal methods of suicide. The restriction of items such as firearms, pesticides, domestic gas, sale of barbiturates, analgesic painkillers; mandatory use of catalytic converters; construction of barriers at jump sites; and lower toxicity antidepressants have been found to be effective in reducing suicide rates. Alcohol Action Ireland called for minimum pricing and other recommendations to reduce alcohol-related harm to be implemented in full.

Sinn Féin's spokesman on health and children, Caoimhghín Ó Caoláin, said the figures for suicide on the island of Ireland were 'truly shocking' and called for a determined and coordinated response from governments on both sides of the border.

Questions

- Poverty is a relative concept. How might young people perceive themselves to be in poverty even if they have a job and a place to live?**
- To what extent would you agree with the view that income inequality is closely linked to other social problems such as those outlined in the article?**
- If income inequality is one of the causes of increased suicide rates in Ireland, how might some of the policies to reduce poverty and inequality outlined in this chapter a) help and b) hinder efforts to reduce suicide rates?**
- Do you think that restricting firearms, pesticides, drugs and other measures outlined along with minimum prices would be more effective in reducing suicides rates than reducing income inequality? Explain.**
- How might an economist identify links between increases in suicide rates and income inequality?**

Family relationship problems can have long-term effects on children and is one of a number of factors put forward to explain high suicide rates amongst males in Ireland.



SUMMARY

- Data on the distribution of income show wide disparity in industrialized economies.
- Because in-kind transfers, the economic life cycle, transitory income and economic mobility are so important for understanding variation in income, it is difficult to gauge the degree of inequality in our society using data on the distribution of income in a single year. When these other factors are taken into account, they tend to suggest that economic well-being is more equally distributed than is annual income.
- Political philosophers differ in their views about the role of government in altering the distribution of income. Utilitarians (such as John Stuart Mill) would choose the distribution of income to maximize the sum of utility of everyone in society. Liberals (such as John Rawls) would determine the distribution of income as if we were behind a ‘veil of ignorance’ that prevented us from knowing our own stations in life. Libertarians (such as Robert Nozick) would have the government enforce individual rights to ensure a fair process but then not be concerned about inequality in the resulting distribution of income.
- Various policies aim to help the poor – minimum wage laws, social security, negative income taxes and in-kind transfers. Although each of these policies helps some families escape poverty, they also have unintended side effects. Because financial assistance declines as income rises, the poor often face effective marginal tax rates that are very high. Such high effective tax rates discourage poor families from escaping poverty on their own.

QUESTIONS FOR REVIEW

- 1 What is meant by the terms ‘income inequality’ and ‘income distribution’?
- 2 Explain the Lorenz curve and what it measures.
- 3 How is the Gini coefficient calculated and what does it show?
- 4 How does the extent of income inequality in your country compare to that of other nations around the world?
- 5 Poverty is described as a relative concept – what does this mean?
- 6 Distinguish between absolute and relative poverty.
- 7 When gauging the amount of inequality, why do transitory and life-cycle variations in income cause difficulties?
- 8 How would a utilitarian, a liberal and a libertarian determine how much income inequality is permissible?
- 9 What are the pros and cons of in-kind (rather than cash) transfers to the poor?
- 10 Describe how anti-poverty programmes can discourage the poor from working. How might you reduce this disincentive? What are the disadvantages with your proposed policy?

PROBLEMS AND APPLICATIONS

- 1 What factors might account for the levels of income inequality that exist in a country? Use the country you are studying in as a case study to help illustrate your answer.
- 2 The Gini coefficient in Portugal, measured on a scale from 0–100, was 34.2 in 2011 whereas the Gini coefficient in Norway was 22.9. Does this mean that the bottom 20 per cent of the population in Portugal must be poorer than the equivalent quintile in Norway? Explain.
- 3 Economists often view life-cycle variation in income as one form of transitory variation in income around people’s lifetime, or permanent, income. In this sense, how does your current income compare to your permanent income? Do you think your current income accurately reflects your standard of living?
- 4 The chapter discusses the importance of economic mobility.
 - a. What policies might the government pursue to increase economic mobility within a generation?
 - b. What policies might the government pursue to increase economic mobility across generations?
 - c. Do you think we should reduce spending on social security benefits in order to increase spending on government programmes that enhance economic mobility? What are some of the advantages and disadvantages of doing so?

- 5** Consider two communities. In one community, ten families have incomes of €100 each and ten families have incomes of €20 each. In the other community, ten families have incomes of €200 each and ten families have incomes of €22 each.
- In which community is the distribution of income more unequal? In which community is the problem of poverty likely to be worse?
 - Which distribution of income would Rawls prefer? Explain.
 - Which distribution of income do you prefer? Explain.
- 6** The chapter uses the analogy of a 'leaky bucket' to explain one constraint on the redistribution of income.
- What elements of your country's system for redistributing income create the leaks in the bucket? Be specific.
 - Do you think that people with left-wing political views believe that the bucket used for redistributing income is more or less leaky than it is believed to be by people with more right-wing political views? How does that belief affect their views about the amount of income redistribution that the government should undertake?
- 7** Suppose there are two possible income distributions in a society of ten people. In the first distribution, nine people would have incomes of €30,000 and one person would have an income of €10,000. In the second distribution, all ten people would have incomes of €25,000.
- If the society had the first income distribution, what would be the utilitarian argument for redistributing income?
 - Which income distribution would Rawls consider more equitable? Explain.
 - Which income distribution would Nozick consider more equitable? Explain.
- 8** Suppose that a family's tax liability equalled its income multiplied by one-half, minus €10,000. Under this system, some families would pay taxes to the government, and some families would receive money from the government through a 'negative income tax'.
- Consider families with pre-tax incomes of €0, €10,000, €20,000, €30,000 and €40,000. Make a table showing pre-tax income, taxes paid to the government or money received from the government, and after-tax income for each family.
 - What is the marginal tax rate in this system (i.e. out of every €1 of extra income, how much is paid in tax)? What is the maximum amount of income at which a family receives money from the government?
 - Now suppose that the tax schedule is changed, so that a family's tax liability equals its income multiplied by one-quarter, minus €10,000. What is the marginal tax rate in this new system? What is the maximum amount of income at which a family receives money from the government?
 - What is the main advantage of each of the tax schedules discussed here?
- 9** John and Carlos are utilitarians. John believes that labour supply is highly elastic, whereas Carlos believes that labour supply is quite inelastic. How do you suppose their views about income redistribution differ?
- 10** Do you agree or disagree with each of the following statements? What do your views imply for public policies, such as taxes on inheritance?
- 'Every parent has the right to work hard and save in order to give his or her children a better life.'
 - 'No child should be disadvantaged by the sloth or bad luck of his or her parents.'

PART 9

TRADE

19 INTERDEPENDENCE AND THE GAINS FROM TRADE

Consider your typical day. You wake up in the morning and you make yourself some coffee from beans grown in Kenya, or tea from leaves grown in Sri Lanka. Over breakfast, you listen to a radio programme on your radio set made in Japan. You get dressed in clothes manufactured in Thailand. You drive to the university in a car made of parts manufactured in more than a dozen countries around the world. Then you open up your economics textbook published by a company located in Hampshire in the UK, printed on paper made from trees grown in Finland and written by authors from the US and England.

Every day you rely on many people from around the world, most of whom you do not know, to provide you with the goods and services that you enjoy. Such interdependence is possible because people trade with one another. Those people who provide you with goods and services are not acting out of generosity or concern for your welfare. Nor is some government or supra-governmental agency directing them to make products you want and to give them to you. Instead, people provide you and other consumers with the goods and services they produce because they get something in return.

THE PRODUCTION POSSIBILITIES FRONTIER

We begin our analysis by looking at a model of the economy where no trade takes place. Although real economies produce thousands of goods and services, this model will simplify the production of goods into two categories, capital goods and consumer goods. The country has a certain amount of resources of land, labour and capital available to allocate between the production of capital goods and consumer goods. The way in which the country can divide the allocation of resources between these two goods can be shown using a production possibilities frontier (PPF). (You might also see this referred to as a production possibilities boundary or production possibilities curve – they are the same thing). The **production possibilities frontier** is a graph that shows the various combinations of output – in this case, capital goods and consumer goods – that the economy can possibly produce given the available factors of production and the available production technology that firms can use to turn these factors into output.

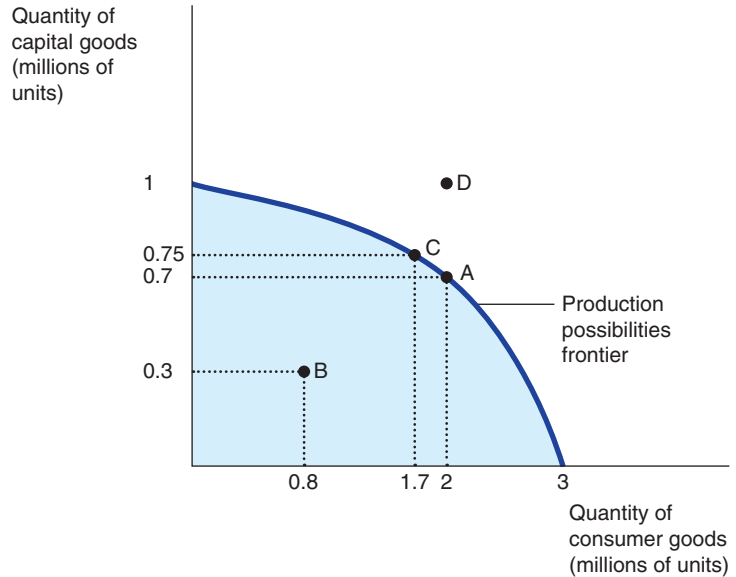
production possibilities frontier a graph that shows the combinations of output that the economy can possibly produce given the available factors of production and available production technology

Figure 19.1 is an example of a PPF. In this economy, if all resources were devoted to the production of capital goods, the economy would produce 1 million units of capital goods and no consumption goods.

FIGURE 19.1

The Production Possibilities Frontier

The production possibilities frontier shows the combinations of output – in this case, capital and consumer goods – that the economy can possibly produce given its factor endowment. The economy can produce any combination on or inside the frontier. Points outside the frontier are not feasible given the economy's resources.



If all resources were used to produce consumer goods, the economy would produce 3 million units and no capital goods. The two end points of the production possibilities frontier represent these extreme possibilities. If the economy were to divide its resources between the two goods, it could produce 700,000 capital goods and 2 million consumer goods, shown in the figure by point A. By contrast, the outcome at point D is not possible because resources are scarce: the economy does not have enough of the factors of production to support that level of output. In other words, the economy can produce at any point on or inside the production possibilities frontier, but it cannot produce at points outside the frontier.

An outcome is said to be *efficient* if the economy is getting all it can from the scarce resources it has available. Points on the PPF represent efficient levels of production. When the economy is producing at such a point, say point A, there is no way to produce more of one good without producing less of the other. Point B represents an *inefficient* outcome. For some reason, the economy is producing less than it could from the resources it has available and so some of its resources are lying idle (they are unemployed or underemployed). At point B, the country is producing only 300,000 units of capital goods and 800,000 consumer goods. If the source of the inefficiency were eliminated, the economy could move from point B to point A, increasing production of both capital goods (to 700,000) and consumer goods (to 2 million).

One of the *Ten Principles of Economics* discussed is that people face trade-offs. The production possibilities frontier shows one trade-off that society faces. Once we have reached the efficient points on the frontier, the only way of getting more of one good is to get less of the other. When the economy moves from point A to point C, for instance, society produces more consumer goods but at the expense of producing fewer consumer goods.

Another of the *Ten Principles of Economics* is that the cost of something is what you give up to get it. This is called the *opportunity cost* (OC). The production possibilities frontier shows the opportunity cost of one good as measured in terms of the other good. When society reallocates some of the factors of production from the capital goods industry to the consumer goods industry, moving the economy from point A to point C, it gives up 300,000 consumer goods to get 50,000 additional capital goods. In other words, when the economy is at point A, the opportunity cost of 300,000 consumer goods is 50,000 capital goods.

Calculating Opportunity Costs

Remember that the opportunity cost is the cost expressed in terms of the next best alternative sacrificed – what has to be given up in order to acquire something. In the example the country had to give up 300,000 consumer goods to acquire 50,000 additional capital goods. The opportunity cost can be expressed in terms of either capital goods or consumer goods – they are the reciprocal of each other.

As a general principle we can express the opportunity cost as a ratio expressed as the sacrifice in one good in terms of the gain in the other:

$$\text{Opportunity cost of good } y = \frac{\text{Sacrifice of good } x}{\text{Gain in good } y}$$

Expressing the opportunity cost in terms of good x would give:

$$\text{Opportunity cost of good } x = \frac{\text{Sacrifice of good } y}{\text{Gain in good } x}$$

What is the opportunity cost of 1 additional unit of consumer goods or capital goods? This can be calculated by the following process. If we want to find the opportunity cost of consumer goods in relation to capital goods then firstly write out the known quantities:

The OC of 300,000 consumer goods is 50,000 capital goods

Now divide both quantities by the number of consumer goods:

$$\text{The OC of } \frac{300,000}{300,000} \text{ consumer goods is } \frac{50,000}{300,000} \text{ capital goods}$$

Now complete the calculation to get the opportunity cost of 1 additional unit of consumer goods in terms of capital goods sacrificed:

The OC of 1 consumer good is 0.16 capital goods

This tells us that for every 1 additional unit of consumer goods acquired, we have to give up 0.16 of a capital good.

To find the opportunity cost of capital goods in terms of consumer goods, follow the same process but in reverse.

The OC of 50,000 capital goods is 300,000 consumer goods

$$\text{The OC of } \frac{50,000}{50,000} \text{ capital goods is } \frac{300,000}{50,000} \text{ consumer goods}$$

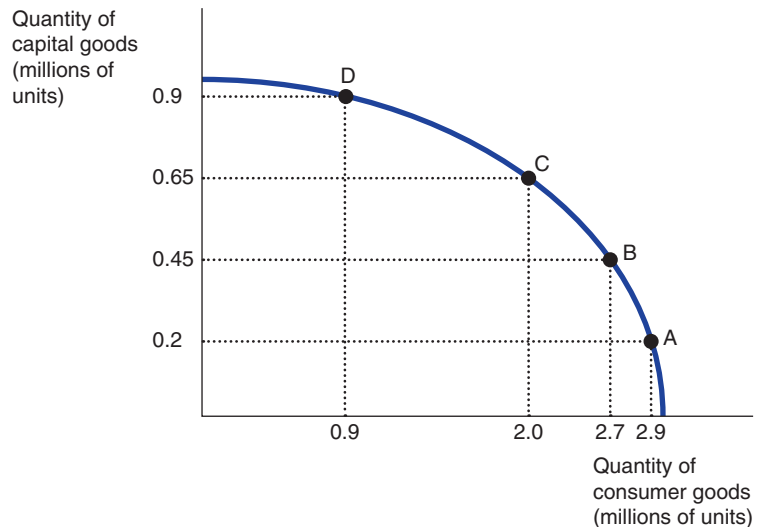
The OC of 1 capital good is 6 consumer goods

The Shape of the Production Possibilities Frontier

The production possibilities frontier in Figure 19.2 is bowed outward (concave to the origin). This is due to the fact that when economies move resources from one use to another, unless they are perfect substitutes (in which case the PPF would be a straight line), as the rate at which the increase in output of one good increases, the opportunity cost in terms of the other changes. For example, if the economy is using most of its resources to make consumer goods at point A in Figure 19.2, land, labour and capital is being used to make consumer goods even if these resources are not best suited to making these goods. If the country moves from point A to point B, the gain in capital goods is 250,000 units but the sacrifice in consumer goods is 200,000. The opportunity cost of 1 additional unit of capital goods will be 0.8 consumer goods sacrificed. Resources released from producing consumer goods are now able to produce capital goods, a use to which they are more appropriately suited. If the economy moves from point B to point C the gain in capital goods is a further 200,000 units but the sacrifice in terms of consumer goods is now 700,000. The opportunity cost of 1 additional unit of capital goods now is 3.5 consumer goods sacrificed.

FIGURE 19.2**The Shape of the Production Possibilities Frontier**

The PPF is concave to the origin. The shape of the PPF reflects the opportunity cost of producing different quantities of capital goods and consumer goods. If the country switches resources from consumer goods to capital goods the opportunity cost in terms of the increase in consumer goods sacrificed for every additional unit of capital goods rises as the output combination moves from point A to point D.



If the country continues to shift resources to capital goods from consumer goods the ease of substitutability of factors becomes weaker and the sacrifice in consumer goods becomes greater. Moving from point C to point D yields an additional 250,000 capital goods but at a cost of 1.1 million units of consumer goods. The opportunity cost of 1 additional unit of capital goods is now 4.4 units of consumer goods. The reason that the opportunity cost in terms of consumer goods is rising is that the resources being put to use producing more capital goods are now less suited to the purpose and so the sacrifice in consumer goods is rising.

The PPF illustrates two of the key questions any economy has to answer: what is to be produced and how will the output be produced? Most economies could use the resources it has at its disposal in a variety of ways. For example, it is possible that the UK could allocate resources to the production of oranges. Large amounts of land could be set aside for the construction of glasshouses in which the climate, water and nutrition needs of orange trees is controlled by computer technology. In Spain the same amount of oranges could be produced using far fewer resources simply because the climate is more conducive to growing oranges. The opportunity cost of using resources in this way in the UK is likely to be high, therefore.

A Shift in the Production Possibilities Frontier

The production possibilities frontier shows the trade-off between the production of different goods at a given time, but the trade-off can change over time. For example, technological advances can mean that factors of production are considerably more productive in terms of output per unit per time period. Countries might also be in a position to recover or make use of more natural resources or the effects of education in the country means the labour force is more productive. Over time, therefore, opportunity cost ratios of production change and this can affect the shape and position of the production possibilities frontier. Figure 19.3 shows three possible outcomes. In panel (a) the PPF has shifted outwards showing that it is now possible to produce more of both capital goods and consumer goods as indicated by the move to point B. If all resources were devoted to capital goods, the economy could now produce 1.2 million units and if all resources were devoted to consumer goods the country could now produce 3.6 million units. The relative opportunity cost ratios, however, remain the same because the PPF has shifted outwards parallel at every point to the original curve. Because of this economic growth, society might move production from point A to point B, enjoying more consumer goods and more capital goods.

Panel (b) shows a shift in the PPF but this time the economic advances in the productivity of the capital goods industry is greater than that of the consumer goods industry. If all resources were now devoted to capital goods production, the country could produce 1.5 million units and if it devoted all output to consumer goods it could produce 3.2 million units. The opportunity cost ratio at all points on the new PPF will now be different to those on the original PPF.

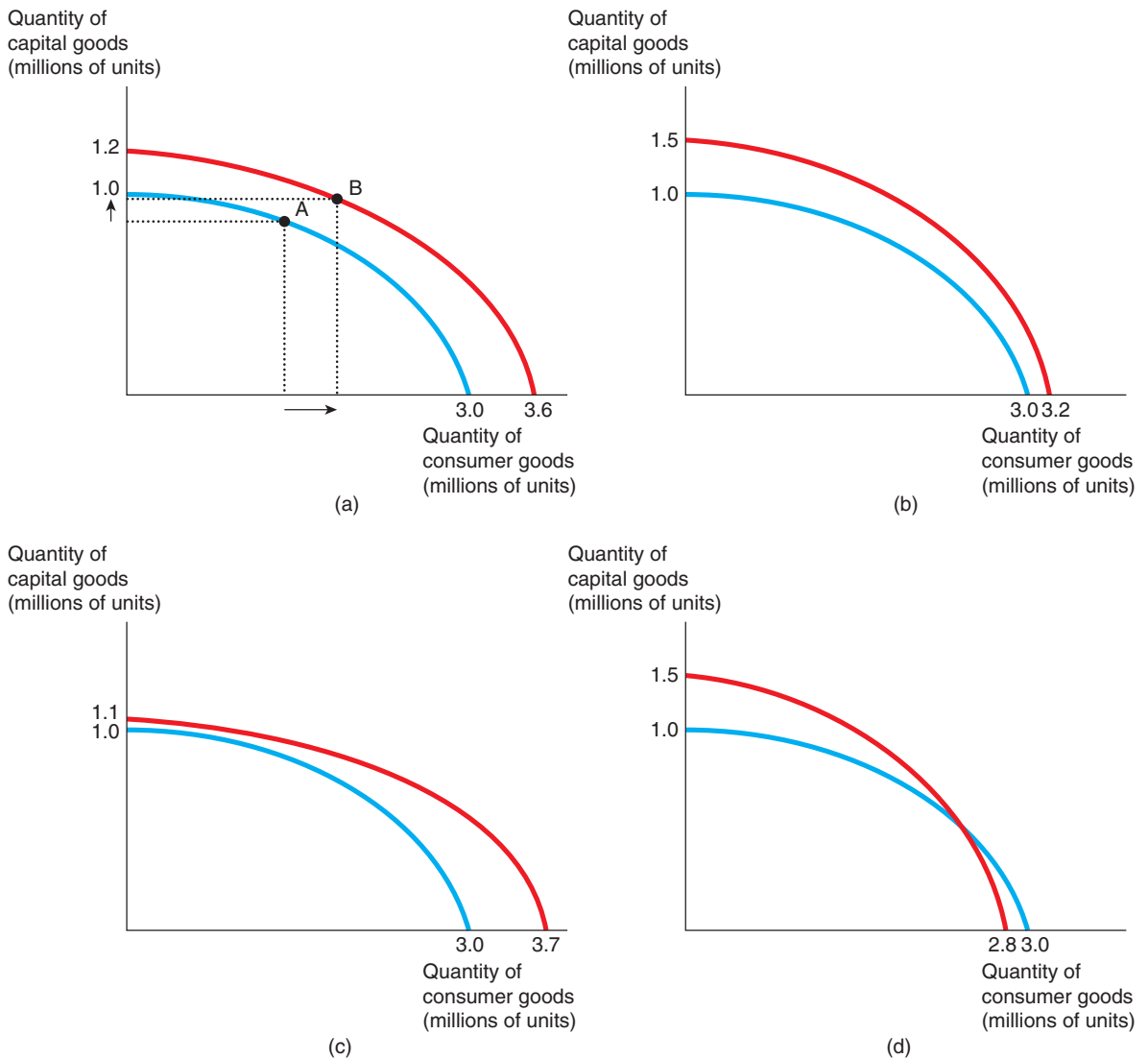
Panel (c) shows the opposite situation where the economic advances in productivity in the consumer goods industry is greater than in the capital goods industry. In this case, if all resources were devoted to producing consumer goods, the country could now produce 3.7 million units and if all resources were devoted to producing capital goods, 1.1 million units could be produced.

It may also be possible that productivity in one industry might actually reduce whilst that in the other rises in which case the PPF might take the look of that in panel (d) of Figure 19.3. In this case there has been an increase in productivity in the capital goods industry but a reduction in the consumer goods industry.

FIGURE 19.3

Shifts in the Production Possibilities Frontier

Panels (a) to (d) show different shifts in the PPF as a result of changes in the productivity of factors used in producing capital and consumer goods.



The production possibilities frontier simplifies a complex economy to highlight and clarify some basic ideas. We can now extend the analysis to look at how different factor endowments and factor productivity in different countries can lead to countries trading and gaining advantages.

SELF TEST Use the information in Figure 19.3 to calculate the opportunity cost ratios of both capital and consumer goods in each scenario presented in the different panels. Assume, in each case, that the country moves from devoting all its resources to capital goods and then switches to devoting all its resources to consumer goods.

INTERNATIONAL TRADE

Each country has its own PPF and in isolation faces choices about the use of resources to produce goods. If the country wants to increase the amount of goods available to all its citizens then it can rely on increasing the number of factors it has at its disposal or the efficiency with which its resources are used to shift the PPF outwards. In addition to this, countries might also choose to engage in trade as a means of providing benefits to its citizens and effectively shift the PPF. One of the *Ten Principles of Economics* is that trade can make everyone better off. This principle explains why people trade with their neighbours and why nations trade with other nations.

A Parable for the Modern Economy

We are going to use a simple example to illustrate how trade can benefit all. Imagine that there are two goods in the world – beef and potatoes. And there are two people in the world (our analogy for two different countries) – a cattle farmer named Silvia and a market gardener named Johan – each of whom would like to eat both beef and potatoes.

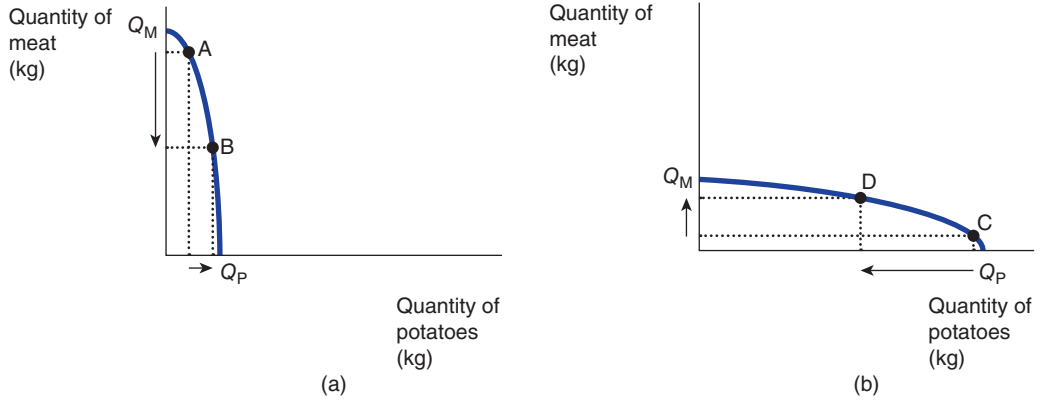
The gains from trade are most obvious if the cattle farmer can produce only meat and the market gardener can produce only potatoes. In one scenario, the farmer and the gardener could choose to have nothing to do with each other. But after several months of eating beef roasted, boiled, fried and grilled, the cattle farmer might decide that self-sufficiency is not all she expected. The market gardener, who has been eating potatoes mashed, fried and baked, would most likely agree. It is easy to see that trade would allow them to enjoy greater variety: each could then have steak and chips.

Although this scene illustrates most simply how everyone can benefit from trade, the gains would be similar if the farmer and the gardener were each capable of producing the other good. In this case each has a PPF analogous to two different countries. Suppose, for example, that the market gardener is able to rear cattle and produce meat, but that he is not very good at it. Similarly, suppose that the cattle farmer is able to grow potatoes, but that her land is not very well suited for it. The PPF for Johan and Silvia would look like those in Figure 19.4. Panel (a) shows Silvia's PPF and panel (b) shows Johan's. Because Silvia is more efficient in the production of meat than she is producing potatoes, the shape of the PPF reflects the opportunity cost of any decision to divert more of her resources away from producing meat to producing potatoes. If Silvia devoted all her time and resources to producing potatoes she could produce Q_M , the vertical intercept. If she makes the decision to divert resources to the production of meat the sacrifice in terms of lost meat output is relatively high compared to the gains in output of potatoes. The opportunity cost of shifting resources from meat to potatoes is high in terms of the sacrifice in meat output she makes to secure a relatively small amount of additional potatoes as shown by the move from point A to point B.

Johan's situation is the reverse of this and is shown in panel (b). If Johan allocates all resources to producing potatoes, he produces Q_P . If he diverts resources away from potato production towards meat production he sacrifices a relatively large amount of output of potatoes to gain a relatively small amount of meat. The opportunity cost of diverting resources to meat for Johan is high, therefore.

FIGURE 19.4**Differing Opportunity Cost Ratios**

Panels (a) and (b) show Silvia and Johan's PPFs respectively. Silvia is more skilled in producing meat whilst Johan is more skilled in producing potatoes, although both could divert resources to produce the other good. The opportunity cost ratios for each are different – the opportunity cost for Silvia of diverting resources from meat to potatoes is high whilst for Johan the opportunity cost of diverting resources from potatoes to meat is high.



Economically, it would be more efficient for Johan and Silvia to cooperate with each other, specialize in what they both do best and benefit from trading with each other at some mutually agreeable rate of exchange. For example, they might agree to a rate of exchange of 1 kg of meat for every 5 kg of potatoes.

The gains from trade are less obvious, however, when one person is better at producing *every* good. For example, suppose that Silvia is better at rearing cattle *and* better at growing potatoes than Johan. In this case, should the farmer or gardener choose to remain self-sufficient, or is there still reason for them to trade with each other? To answer this question, we need to look more closely at the factors that affect such a decision.

Production Possibilities

Suppose that Johan and Silvia each work 8 hours a day 6 days a week (a working week of 48 hours) and take Sunday off. They can spend their time growing potatoes, rearing cattle, or a combination of the two. Table 19.1 shows the amount of time each person takes to produce 1 kg of each good. The gardener can produce 1 kg of meat in 6 hours and 1 kg of potatoes in an hour and a half. The farmer, who is more productive in both activities, can produce a kilogram of meat in 2 hours and a kilogram of potatoes in 1 hour. The last columns in Table 19.1 show the amounts of meat or potatoes the gardener and farmer can produce in a 48-hour working week, producing only that good.

TABLE 19.1**The Production Opportunities of Johan the Gardener and Silvia the Farmer**

| | Time needed to make 1 kg of: | | Amount of meat or potatoes produced in 48 hours | |
|--------|------------------------------|------------|---|----------|
| | Meat | Potatoes | Meat | Potatoes |
| Johan | 6 hrs/kg | 1.5 hrs/kg | 8 kg | 32 kg |
| Silvia | 2 hrs/kg | 1 hr/kg | 24 kg | 48 kg |

Panel (a) of Figure 19.5 illustrates the amounts of meat and potatoes that Johan can produce. If he devotes all 48 hours of his time to potatoes, he produces 32 kg of potatoes (measured on the horizontal axis) and no meat. If he devotes all his time to meat, he produces 8 kg of meat (measured on the vertical axis) and no potatoes. If Johan divides his time equally between the two activities, spending 24 hours a week on each, he produces 16 kg of potatoes and 4 kg of meat. The figure shows these three possible outcomes and all others in between.

FIGURE 19.5

The Production Possibilities Frontier

Panel (a) shows the combinations of meat and potatoes that the gardener can produce. Panel (b) shows the combinations of meat and potatoes that the farmer can produce. Both production possibilities frontiers are derived from Table 19.1 and the assumption that the gardener and farmer each work 8 hours a day.



This graph is Johan's production possibilities frontier. Note that in this case, the PPF is a straight line indicating that the slope is constant and thus the opportunity cost to Johan of switching between potatoes and meat is constant. Johan faces a trade-off between producing meat and producing potatoes. If Johan devotes an extra hour to producing meat he sacrifices potato production. Assume Johan starts at point A producing 4 kg of meat and 16 kg of potatoes. If he then devoted all resources to producing potatoes moving to point B, he would produce 32 kg of potatoes and sacrifice 4 kg of potatoes. The opportunity cost of 1 additional kilo of potatoes, therefore, is 0.25 kg of meat. Every additional 1 kg of potatoes produced would involve a trade-off of $\frac{1}{4}$ kg of meat. If Johan chose to increase meat production by 1 kg he would have to sacrifice 4 kg of potatoes.

Panel (b) of Figure 19.5 shows the production possibilities frontier for Silvia. If she devotes all 48 hours of her working week to potatoes, she produces 48 kg of potatoes and no meat. If she devotes all of her time to meat production, she produces 24 kg of meat and no potatoes. If Silvia divides her time equally, spending 24 hours a week on each activity, she produces 24 kg of potatoes and 12 kg of meat. Once again, the production possibilities frontier shows all the possible outcomes. If Silvia moved from devoting half her time producing each to producing all potatoes, the opportunity cost of the additional 24 kg of potatoes is 12 kg of meat. Silvia would sacrifice $\frac{1}{2}$ kg of meat for every 1 kg of additional potatoes. The slope of this production possibilities frontier is, therefore, 0.5. If Silvia shifted production to meat from potatoes, the opportunity cost of an additional 1 kg of meat would be 2 kg of potatoes sacrificed.

If the gardener and farmer choose to be self-sufficient, rather than trade with each other, then each consumes exactly what he or she produces. In this case, the production possibilities frontier is also the

consumption possibilities frontier. That is, without trade, Figure 19.5 shows the possible combinations of meat and potatoes that Johan and Silvia can each consume.

Although these production possibilities frontiers are useful in showing the trade-offs that the gardener and farmer face, they do not tell us what Johan and Silvia will actually choose to do. To determine their choices, we need to know the tastes of the gardener and the farmer. Let's suppose they choose the combinations identified by points A and B in Figure 19.5: Johan produces and consumes 16 kg of potatoes and 4 kg of meat, while Silvia produces and consumes 24 kg of potatoes and 12 kg of meat.

Specialization and Trade

After several years of feeding her family on combination B, Silvia gets an idea and she goes to talk to Johan:

FARMER: Johan, have I got a deal for you! I know how to improve life for both of us. I think you should stop producing meat altogether and devote all your time to growing potatoes. According to my calculations, if you devote all of your working week to growing potatoes, you'll produce 32 kg of potatoes. If you give me 15 of those 32 kg, I'll give you 5 kg of meat in return. You will have 17 kg of potatoes left to enjoy and also 5 kg of meat every week, instead of the 16 kg of potatoes and 4 kg of meat you now make do with. If you go along with my plan, you'll have more of *both* foods. [To illustrate her point, Silvia shows Johan panel (a) of Figure 19.6.]

GARDENER: (*sounding sceptical*) That seems like a good deal for me, Silvia, but I don't understand why you are offering it. If the deal is so good for me, it can't be good for you too.

FARMER: Oh, but it is, Johan! Suppose I spend 12 hours a week growing potatoes and 36 hours rearing cattle. Then I can produce 12 kg of potatoes and 18 kg of meat. You will give me 15 kg of your potatoes in exchange for the 5 kg of my meat. This means I end up with 27 kg of potatoes and 13 kg of meat. So I will also be able to consume more of both foods than I do now. [She points out panel (b) of Figure 19.6.]

FIGURE 19.6

How Trade Expands the Set of Consumption Opportunities

The proposed trade between the gardener and the farmer offers each of them a combination of meat and potatoes that would be impossible in the absence of trade. In panel (a), the gardener consumes at point A* rather than point A. In panel (b), the farmer consumes at point B* rather than point B. Trade allows each to consume more meat and more potatoes.

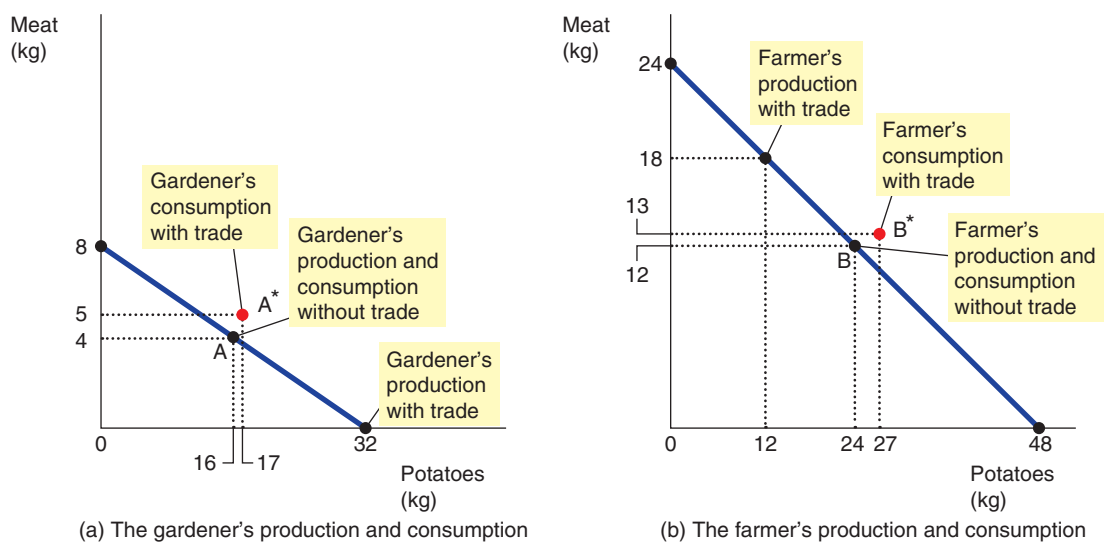


TABLE 19.2 The Gains from Trade: A Summary

| | Gardener | | Farmer | |
|----------------------------|-----------|-------------|------------|------------|
| | Meat | Potatoes | Meat | Potatoes |
| Without trade: | | | | |
| Production and consumption | 4 kg | 16 kg | 12 kg | 24 kg |
| With trade: | | | | |
| Production | 0 kg | 32 kg | 18 kg | 12 kg |
| Trade | Gets 5 kg | Gives 15 kg | Gives 5 kg | Gets 15 kg |
| Consumption | 5 kg | 17 kg | 13 kg | 27 kg |
| Gains from trade: | | | | |
| Increase in consumption | +1 kg | +1 kg | +1 kg | +3 kg |

GARDENER: I don't know, Silvia. ... This sounds too good to be true.

FARMER: It's really not as complicated as it seems at first. Here – I've summarized my proposal for you in a simple table. [The farmer hands the gardener a copy of Table 19.2.]

GARDENER: (*after pausing to study the table*) These calculations seem correct, but I am puzzled. How can this deal make us both better off?

FARMER: We can both benefit because trade allows each of us to specialize in doing what we do best. You will spend more time growing potatoes and less time rearing cattle. I will spend more time rearing cattle and less time growing potatoes. As a result of specialization and trade, each of us can consume more meat and more potatoes without working any more hours.

SELF TEST Draw an example of a production possibilities frontier for Jan, who is stranded on an island after a shipwreck and spends his time gathering coconuts and catching fish. Does this frontier limit Jan's consumption of coconuts and fish if he lives by himself? Does he face the same limits if he can trade with natives on the island?

THE PRINCIPLE OF COMPARATIVE ADVANTAGE

The farmer's explanation of the gains from trade, though correct, poses a puzzle: if the farmer is better at both rearing cattle and growing potatoes, how can the gardener ever specialize in doing what he does best? The gardener doesn't seem to do anything best. To solve this puzzle, we need to look at the principle of *comparative advantage*.

As a first step in developing this principle, consider the following question: in our example, who can produce potatoes at lower cost – the gardener or the farmer? There are two possible answers, and in these two answers lie the solution to our puzzle and the key to understanding the gains from trade. The slope of the production possibilities frontier discussed above will help us to solve the puzzle.

Absolute Advantage

One way to answer the question about the cost of producing potatoes is to compare the inputs required by the two producers. Economists use the term **absolute advantage** when comparing the productivity of one person, firm or nation to that of another. The producer that requires a smaller quantity of inputs to produce a good is said to have an absolute advantage in producing that good.

absolute advantage absolute advantage exists where a producer can produce a good using fewer factor inputs than another

In our example, the farmer has an absolute advantage both in producing meat and in producing potatoes, because she requires less time than the gardener to produce a unit of either good. The farmer needs to input only 2 hours in order to produce a kilogram of meat, whereas the gardener needs 6 hours. Similarly, Silvia needs only 1 hour to produce a kilogram of potatoes, whereas Johan needs 1.5 hours. Based on this information, we can conclude that the farmer has the lower cost of producing potatoes, if we measure cost in terms of the quantity of inputs.

Opportunity Cost and Comparative Advantage

There is another way to look at the cost of producing potatoes. Rather than comparing inputs required, we can compare the opportunity costs. Let's first consider Silvia's opportunity cost in relation to the amount of hours she needs to work. According to Table 19.1, producing 1 kg of potatoes takes her 1 hour of work. When Silvia spends that 1 hour producing potatoes, she spends 1 hour less producing meat. Because Silvia needs 2 hours to produce 1 kg of meat, 1 hour of work would yield $\frac{1}{2}$ kg of meat. Hence, the farmer's opportunity cost of producing 1 kg of potatoes is $\frac{1}{2}$ kg of meat.

Now consider Johan's opportunity cost. Producing 1 kg of potatoes takes him $1\frac{1}{2}$ hours. Because he needs 6 hours to produce 1 kg of meat, $1\frac{1}{2}$ hours of work would yield $\frac{1}{4}$ kg of meat. Hence, the gardener's opportunity cost of 1 kg of potatoes is $\frac{1}{4}$ kg of meat.

Table 19.3 shows the opportunity costs of meat and potatoes for the two producers. Remember that the opportunity cost of meat is the inverse of the opportunity cost of potatoes. Because 1 kg of potatoes costs the farmer $\frac{1}{2}$ kg of meat, 1 kg of meat costs the farmer 2 kg of potatoes. Similarly, because 1 kg of potatoes costs the gardener $\frac{1}{4}$ kg of meat, 1 kg of meat costs the gardener 4 kg of potatoes.

TABLE 19.3

The Opportunity Cost of Meat and Potatoes

| | Opportunity cost of: | |
|----------|----------------------|------------------|
| | 1 kg of meat | 1 kg of potatoes |
| Gardener | 4 kg potatoes | 0.25 kg meat |
| Farmer | 2 kg potatoes | 0.5 kg meat |

Economists use the term **comparative advantage** when describing the opportunity cost of two producers. The producer who gives up less of other goods to produce good X has the smaller opportunity cost of producing good X and is said to have a comparative advantage in producing it.

comparative advantage the comparison among producers of a good according to their opportunity cost. A producer is said to have a comparative advantage in the production of a good if the opportunity cost is lower than that of another producer

In our example, the gardener has a lower opportunity cost of producing potatoes than does the farmer: a kilogram of potatoes costs the gardener only $\frac{1}{4}$ kg of meat, while it costs the farmer $\frac{1}{2}$ kg of meat. Conversely, Silvia has a lower opportunity cost of producing meat than does Johan: a kilogram of meat costs Silvia 2 kg of potatoes, while it costs Johan 4 kg of potatoes. Thus, the gardener has a comparative advantage in growing potatoes, and the farmer has a comparative advantage in producing meat.

Although it is possible for one person to have an absolute advantage in both goods (as Silvia does in our example), it is impossible for one person to have a comparative advantage in both goods. Because the opportunity cost of one good is the inverse of the opportunity cost of the other, if a person's opportunity cost of one good is relatively high, his opportunity cost of the other good must be relatively low. Comparative advantage reflects the relative opportunity cost. Unless two people have exactly the same opportunity cost, one person will have a comparative advantage in one good, and the other person will have a comparative advantage in the other good.

FYI



The Legacy of Adam Smith and David Ricardo

Economists have long understood the principle of comparative advantage. Here is how the great economist Adam Smith put the argument:

It is the maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The tailor does not attempt to make his own shoes, but buys them of the shoemaker. The shoemaker does not attempt to make his own clothes but employs a tailor. The gardener attempts to make neither the one nor the other, but employs those different artificers. All of them find it for their interest to employ their whole industry in a way in which they have some advantage over their neighbours, and to purchase with a part of its produce, or what is the same thing, with the

price of part of it, whatever else they have occasion for.

This quotation is from Smith's 1776 book *An Inquiry into the Nature and Causes of the Wealth of Nations*, which was a landmark in the analysis of trade and economic interdependence.

Smith's book inspired David Ricardo – an Englishman born of Dutch parents – to become an economist, having already made his fortune as a stockbroker in the City of London. In his 1817 book *Principles of Political Economy and Taxation*, Ricardo developed the principle of comparative advantage as we know it today. The Principle was originally put forward by Robert Torrens, a British Army officer and owner of the *Globe* newspaper in 1815. Ricardo's

defence of free trade was not a mere academic exercise. Ricardo put his economic beliefs to work as a member of the British parliament, where he opposed the Corn Laws, which restricted the import of grain.

The conclusions of Adam Smith and David Ricardo on the gains from trade have held up well over time. Although economists often disagree on questions of policy, they are united in their support of free trade. Moreover, the central argument for free trade has not changed much in the past two centuries. Even though the field of economics has broadened its scope and refined its theories since the time of Smith and Ricardo, economists' opposition to trade restrictions is still based largely on the principle of comparative advantage.

Comparative Advantage and Trade

Differences in opportunity cost and comparative advantage create the gains from trade. When each person specializes in producing the good for which he or she has a comparative advantage, total production in the economy rises, and this increase in the size of the economic cake can be used to make everyone better off. In other words, as long as two people have different opportunity costs, each can benefit from trade by obtaining a good at a price that is lower than his or her opportunity cost of that good.

Consider the proposed deal from the viewpoint of Johan. He gets 5 kg of meat in exchange for 15 kg of potatoes. In other words, Johan buys each kilogram of meat for a price of 3 kg of potatoes. This price of meat is lower than his opportunity cost for 1 kg of meat, which is 4 kg of potatoes. Thus, the gardener benefits from the deal because he gets to buy meat at a good price.

Now consider the deal from Silvia's viewpoint. The farmer buys 15 kg of potatoes for a price of 5 kg of meat. That is, the price of potatoes is $\frac{1}{3}$ kg of meat. This price of potatoes is lower than her opportunity cost of 1 kg of potatoes, which is $\frac{1}{2}$ kg of meat. The farmer benefits because she is able to buy potatoes at a good price.

These benefits arise because each person concentrates on the activity for which he or she has the lower opportunity cost: the gardener spends more time growing potatoes, and the farmer spends more time producing meat. As a result, the total production of potatoes and the total production of meat both rise. In our example, potato production rises from 40 to 44 kg, and meat production rises from 16 to 18 kg. The gardener and farmer share the benefits of this increased production. The moral of the

story of the gardener and the farmer should now be clear: *trade can benefit everyone in society because it allows people to specialize in activities in which they have a comparative advantage.*

SELF TEST Jan can gather ten coconuts or catch one fish per hour. His friend, Marie, can gather 30 coconuts or catch two fish per hour. What is Jan's opportunity cost of catching one fish? What is Marie's? Who has an absolute advantage in catching fish? Who has a comparative advantage in catching fish?

The principle of comparative advantage explains interdependence and the gains from trade. Because interdependence is so prevalent in the modern world, the principle of comparative advantage has many applications.

CASE STUDY China's Changing Comparative Advantage

One of the main reasons said to account for the dramatic rise in China as a world trading power in recent years has been its plentiful supply of relatively cheap labour which has helped the country to be able to attain economic growth on the back of significant export volumes into countries across the world.

This source of comparative advantage over other manufacturing countries may well be starting to change. As more and more workers have become employed in manufacturing, the wages they are earning compared to those that existed in primarily agricultural employment are higher, and millions of Chinese workers have enjoyed an improvement in living standards in the last 30 years. China knows that it still faces huge challenges in bringing more workers out of poverty but as living standards rise, so do expectations. Many Chinese workers are now able to access the media and in doing so have been able to see what living standards workers in other developed countries experience. As awareness and expectations increase, so do the demands for improved terms of employment. The number of workers now willing to work for low wages is starting to shrink and at some point this group will disappear. At the same time those in work will expect to see their wages increasing and there are already signs that this is the case. If wages increase at a faster rate than those in other low-wage manufacturing economies, China's comparative advantage in low labour costs will diminish.

To maintain the economic growth that the country has witnessed in the last 30 years, therefore, observers of China are suggesting that the country will need to maintain its comparative advantage by investing in science and technology and innovation so that it can shift its productive emphasis from low value-added manufactured products to higher value-added production. If it does so then it will be able to develop a new comparative advantage and this will provide the impetus for a further period of sustained economic growth.



A country's comparative advantage changes over time. In the case of China, the comparative advantage it enjoyed through having a plentiful supply of cheap labour is beginning to diminish as it faces competition from other low wage nations.

Should Countries in Europe Trade with other Countries?

Just as individuals can benefit from specialization and trade with one another, as the gardener and farmer did, so can populations of people in different countries (remember Johan and Silvia were our analogies for countries). Many of the goods that Europeans enjoy are produced abroad, and many of the goods produced across Europe are sold abroad. Goods produced abroad and purchased for use in the domestic economy are called **imports**. An import leads to a flow of money from the country in payment. Goods produced domestically and sold abroad are called **exports**.

imports goods produced abroad and purchased for use in the domestic economy leading to an outflow of funds from a country

exports goods produced domestically and sold abroad leading to an inflow of funds into a country

To see how countries can benefit from trade, suppose there are two countries, Germany and the Netherlands, and two goods, machine tools and cut flowers. Imagine that the two countries produce cut flowers equally well: a German worker and a Dutch worker can each produce 1 tonne per month. By contrast, because Germany has more land suitable for manufacturing, it is better at producing machine tools: a German worker can produce 2 tonnes of machine tools per month, whereas a Dutch worker can produce only 1 tonne of machine tools per month.

The principle of comparative advantage states that each good should be produced by the country that has the smaller opportunity cost of producing that good. Because the opportunity cost of an additional 1 tonne of cut flowers is 2 tonnes of machine tools in Germany but only 1 tonne of machine tools in the Netherlands, the Dutch have a comparative advantage in producing cut flowers. The Netherlands should produce more cut flowers than it wants for its own use and export some of them to Germany. Similarly, because the opportunity cost of a tonne of cut flowers is 1 tonne of machine tools in the Netherlands but only $\frac{1}{2}$ a car in Germany, the Germans have a comparative advantage in producing machine tools. Germany should produce more machine tools than it wants to consume and export some of it to the Netherlands. Through specialization and trade, both countries can have more machine tools and more cut flowers.

In reality, of course, the issues involved in trade among nations are more complex than this simple example suggests, as we will see later. Most important among these issues is that each country has many citizens with different interests. International trade can make some individuals worse off, even as it makes the country as a whole better off. When Germany exports machine tools and imports cut flowers, the impact on a German worker is not the same as the impact on a German cut flower worker. Yet, contrary to the opinions sometimes voiced by politicians and political commentators, international trade is not like war, in which some countries win and others lose. Trade allows all countries to achieve greater prosperity.

SELF TEST Suppose that the world's fastest typist happens to be trained in brain surgery. Should he do his own typing or hire a secretary? Explain.

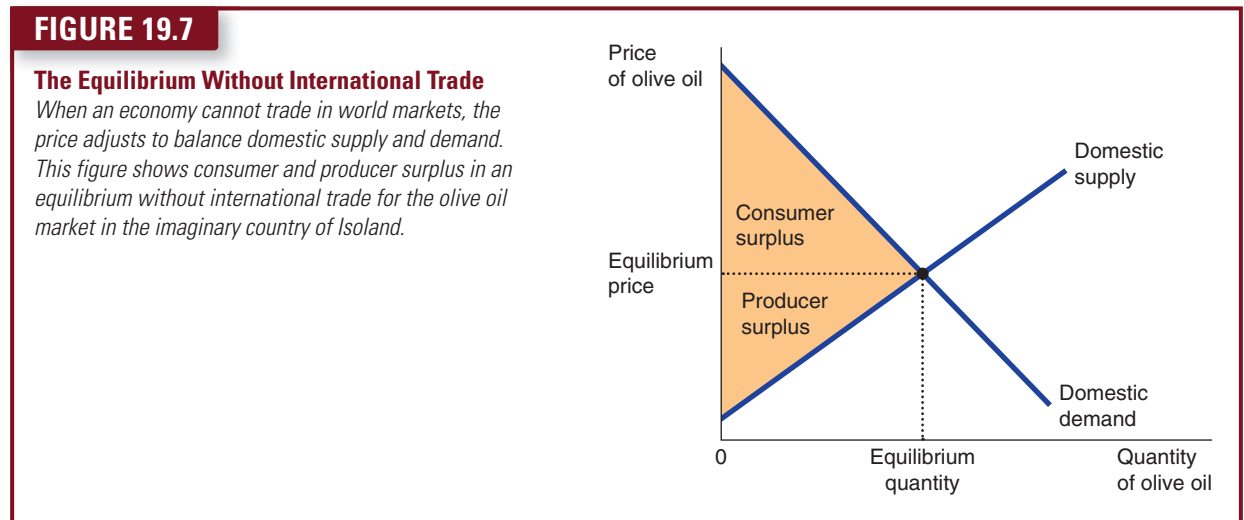
THE DETERMINANTS OF TRADE

Having seen that there are benefits to countries of trading, in this next section we look at the gains and losses of international trade. Consider the market for olive oil. The olive oil market is well suited to examining the gains and losses from international trade: olive oil is made in many countries around the world, and there is much world trade in olive oil. Moreover, the olive oil market is one in which policymakers often consider (and sometimes implement) trade restrictions to protect domestic olive oil producers from foreign competitors. We examine here the olive oil market in the imaginary country of Isoland.

The Equilibrium Without Trade

Assume that the Isolandian olive oil market is isolated from the rest of the world. By government decree, no one in Isoland is allowed to import or export olive oil, and the penalty for violating the decree is so large that no one dares try.

Because there is no international trade, the market for olive oil in Isoland consists solely of Isolandian buyers and sellers. As Figure 19.7 shows, the domestic price adjusts to balance the quantity supplied by domestic sellers and the quantity demanded by domestic buyers. The figure shows the consumer and producer surplus in the equilibrium without trade. The sum of consumer and producer surplus measures the total benefits that buyers and sellers receive from the olive oil market.



Now suppose that Isoland elects a new president. The president campaigned on a platform of ‘change’ and promised the voters bold new ideas. Her first act is to assemble a team of economists to evaluate Isolandian trade policy. She asks them to report back on three questions:

- If the government allowed Isolandians to import and export olive oil, what would happen to the price of olive oil and the quantity of olive oil sold in the domestic olive oil market?
- Who would gain from free trade in olive oil and who would lose, and would the gains exceed the losses?
- Should a tariff (a tax on olive oil imports) or an import quota (a limit on olive oil imports) be part of the new trade policy?

The World Price and Comparative Advantage

The first issue our economists take up is whether Isoland is likely to become an olive oil importer or an olive oil exporter. In other words, if free trade were allowed, would Isolandians end up buying or selling olive oil in world markets? To answer this question, the economists compare the current Isolandian price of olive oil with the price of olive oil in other countries. We call the price prevailing in world markets the **world price**. If the world price of olive oil is higher than the domestic price, then Isoland would become an exporter of olive oil once trade is permitted. Isolandian olive oil producers would be eager to receive the higher prices available abroad and would start selling their olive oil to buyers in other countries. Conversely, if the world price of olive oil is lower than the domestic price, then Isoland would become an importer of olive oil. Because foreign sellers offer a better price, Isolandian olive oil consumers would quickly start buying olive oil from other countries.

world price the price of a good that prevails in the world market for that good

In essence, comparing the world price and the domestic price before trade indicates whether Isoland has a comparative advantage in producing olive oil. The domestic price reflects the opportunity cost of olive oil: it tells us how much an Isolandian must give up to get one unit of olive oil. If the domestic price is low, the cost of producing olive oil in Isoland is low, suggesting that Isoland has a comparative advantage in producing olive oil relative to the rest of the world. If the domestic price is high, then the cost of producing olive oil in Isoland is high, suggesting that foreign countries have a comparative advantage in producing olive oil. By comparing the world price and the domestic price before trade, we can determine whether Isoland is better or worse at producing olive oil than the rest of the world.

SELF TEST The country Autarka does not allow international trade. In Autarka, you can buy a wool suit for 300 grams of gold. Meanwhile, in neighbouring countries, you can buy the same suit for 200 grams of gold. If Autarka were to allow free trade, would it import or export suits? Why?

THE WINNERS AND LOSERS FROM TRADE

To analyse the welfare effects of free trade, the Isolandian economists begin with the assumption that Isoland is a small economy compared with the rest of the world so that its actions have a negligible effect on world markets. The small economy assumption has a specific implication for analysing the olive oil market: if Isoland is a small economy, then the change in Isoland's trade policy will not affect the world price of olive oil. The Isolandians are said to be *price takers* in the world economy. That is, they take the world price of olive oil as given. They can sell olive oil at this price and be exporters, or buy olive oil at this price and be importers.

The small economy assumption is not necessary to analyse the gains and losses from international trade. But the Isolandian economists know from experience that this assumption greatly simplifies the analysis. They also know that the basic lessons do not change in the more complicated case of a large economy.

The Gains and Losses of an Exporting Country

Figure 19.8 shows the Isolandian olive oil market when the domestic equilibrium price before trade is below the world price. Once free trade is allowed, the domestic price rises to equal the world price. No seller of olive oil would accept less than the world price, and no buyer would pay more than the world price.

With the domestic price now equal to the world price, the domestic quantity supplied differs from the domestic quantity demanded. The supply curve shows the quantity of olive oil supplied by Isolandian sellers. The demand curve shows the quantity of olive oil demanded by Isolandian buyers. Because the domestic quantity supplied is greater than the domestic quantity demanded, Isoland sells olive oil to other countries. Thus, Isoland becomes an olive oil exporter.

Although domestic quantity supplied and domestic quantity demanded differ, the olive oil market is still in equilibrium because there is now another participant in the market: the rest of the world. One can view the horizontal line at the world price as representing the demand for olive oil from the rest of the world. This demand curve is perfectly elastic because Isoland, as a small economy, can sell as much olive oil as it wants at the world price.

Now consider the gains and losses from opening up trade. Clearly, not everyone benefits. Trade forces the domestic price to rise to the world price. Domestic producers of olive oil are better off because they can now sell olive oil at a higher price, but domestic consumers of olive oil are worse off because they have to buy olive oil at a higher price.

To measure these gains and losses, we look at the changes in consumer and producer surplus, which are shown in the graph and table in Figure 19.9. Before trade is allowed, the price of olive oil adjusts to balance domestic supply and domestic demand. Consumer surplus, the area between the demand curve

FIGURE 19.8

International Trade in an Exporting Country

Once trade is allowed, the domestic price rises to equal the world price. The supply curve shows the quantity of olive oil produced domestically, and the demand curve shows the quantity consumed domestically. Exports from Isoland equal the difference between the domestic quantity supplied and the domestic quantity demanded at the world price.

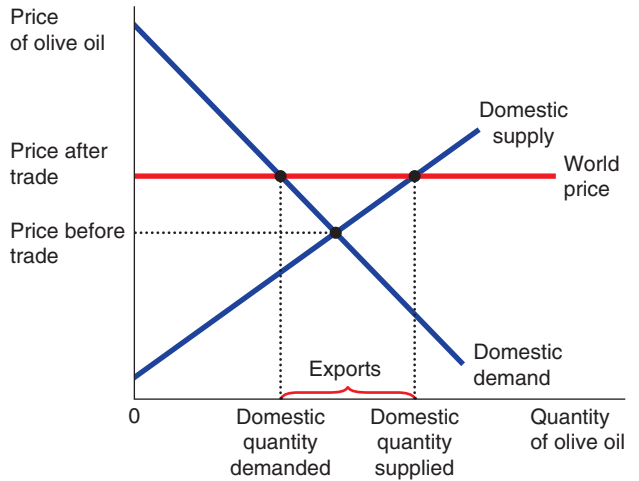


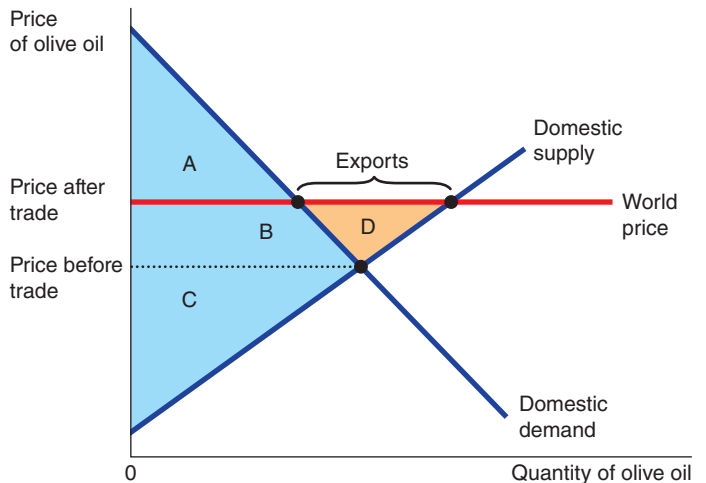
FIGURE 19.9

How Free Trade Affects Welfare in an Exporting Country

When the domestic price rises to equal the world price, sellers are better off (producer surplus rises from C to $B + C + D$), and buyers are worse off (consumer surplus falls from $A + B$ to A). Total surplus rises by an amount equal to area D , indicating that trade raises the economic well-being of the country as a whole.

| | Before trade | After trade | Change |
|------------------|--------------|-----------------|------------|
| Consumer surplus | $A + B$ | A | $-B$ |
| Producer surplus | C | $B + C + D$ | $+(B + D)$ |
| Total surplus | $A + B + C$ | $A + B + C + D$ | $+D$ |

The area D shows the increase in total surplus and represents the gains from trade.



and the before-trade price, is area $A + B$. Producer surplus, the area between the supply curve and the before-trade price, is area C . Total surplus before trade, the sum of consumer and producer surplus, is area $A + B + C$.

After trade is allowed, the domestic price rises to the world price. Consumer surplus is area A (the area between the demand curve and the world price). Producer surplus is area $B + C + D$ (the area between the supply curve and the world price). Thus, total surplus with trade is area $A + B + C + D$.

These welfare calculations show who wins and who loses from trade in an exporting country. Sellers benefit because producer surplus increases by the area $B + D$. Buyers are worse off because consumer surplus decreases by the area B . Because the gains of sellers exceed the losses of buyers by the area D , total surplus in Isoland increases.

This analysis of an exporting country yields two conclusions:

- When a country allows trade and becomes an exporter of a good, domestic producers of the good are better off and domestic consumers of the good are worse off.
- Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.

The Gains and Losses of an Importing Country

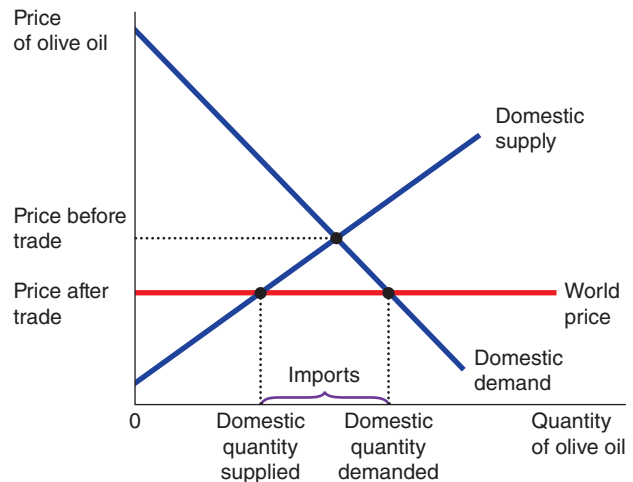
Now suppose that the domestic price before trade is above the world price. Once again, after free trade is allowed, the domestic price must equal the world price. As Figure 19.10 shows, the domestic quantity supplied is less than the domestic quantity demanded. The difference between the domestic quantity demanded and the domestic quantity supplied is bought from other countries, and Isoland becomes an olive oil importer.

In this case, the horizontal line at the world price represents the supply of the rest of the world. This supply curve is perfectly elastic because Isoland is a small economy and, therefore, can buy as much olive oil as it wants at the world price.

FIGURE 19.10

International Trade in an Importing Country

Once trade is allowed, the domestic price falls to equal the world price. The supply curve shows the amount produced domestically, and the demand curve shows the amount consumed domestically. Imports equal the difference between the domestic quantity demanded and the domestic quantity supplied at the world price.



Now consider the gains and losses from trade. Once again, not everyone benefits. When trade forces the domestic price to fall, domestic consumers are better off (they can now buy olive oil at a lower price), and domestic producers are worse off (they now have to sell olive oil at a lower price). Changes in consumer and producer surplus measure the size of the gains and losses, as shown in the graph and table in Figure 19.11. Before trade, consumer surplus is area A, producer surplus is area B + C, and total surplus is area A + B + C. After trade is allowed, consumer surplus is area A + B + D, producer surplus is area C and total surplus is area A + B + C + D.

These welfare calculations show who wins and who loses from trade in an importing country. Buyers benefit because consumer surplus increases by the area B + D. Sellers are worse off because producer surplus falls by the area B. The gains of buyers exceed the losses of sellers, and total surplus increases by the area D.

This analysis of an importing country yields two conclusions parallel to those for an exporting country:

- When a country allows trade and becomes an importer of good, domestic consumers of the good are better off and domestic producers of the good are worse off.
- Trade raises the economic well-being of a nation in the sense that the gains of the winners exceed the losses of the losers.

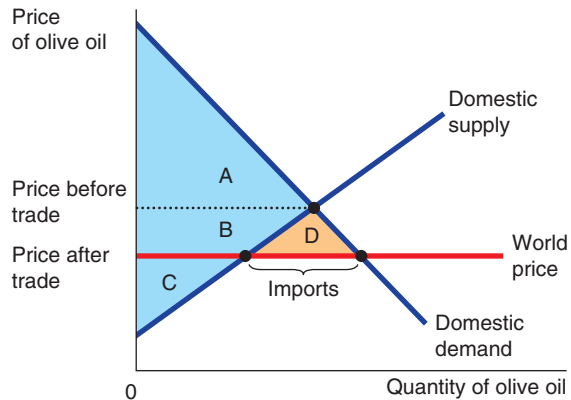
FIGURE 19.11

How Free Trade Affects Welfare in an Importing Country

When the domestic price falls to equal the world price, buyers are better off (consumer surplus rises from A to $A + B + D$), and sellers are worse off (producer surplus falls from $B + C$ to C). Total surplus rises by an amount equal to area D , indicating that trade raises the economic well-being of the country as a whole.

| | Before trade | After trade | Change |
|------------------|--------------|-----------------|------------|
| Consumer surplus | A | $A + B + D$ | $+(B + D)$ |
| Producer surplus | $B + C$ | C | $-B$ |
| Total surplus | $A + B + C$ | $A + B + C + D$ | $+D$ |

The area D shows the increase in total surplus and represents the gains from trade.



Having completed our analysis of trade, we can better understand one of the *Ten Principles of Economics*: trade can make everyone better off. If Isoland opens up its olive oil market to international trade the change will create winners and losers, regardless of whether Isoland ends up exporting or importing olive oil. Notice that in our analysis, we have not made a judgement about the winners and losers – whether the gain to the producers is more valuable than the loss to the consumers. The important issue, from an economic analysis perspective, is the effect on total welfare, which in this case has risen for Isoland. In the real world, policymakers may have to take into consideration the power which resides with different groups. If domestic consumers of olive oil in Isoland had considerable lobbying power compared with olive oil producers then policy decisions may be affected which distort outcomes and reduce total welfare. The effect on consumers, for example, might be limited in comparison with the gains to producers but presenting arguments in this way does not always win political points! This is something that must always be considered because whilst economic analysis may point to a clear policy decision and outcome, there are many other factors that decision makers have to take into account, as exemplified when we look at the arguments for restricting trade.

In our example, the gains of the winners exceed the losses of the losers, so the winners could compensate the losers and still be better off. In this sense, trade *can* make everyone better off. But *will* trade make everyone better off? Probably not. In practice, compensation for the losers from international trade is rare. Without such compensation, opening up to international trade is a policy that expands the size of the economic cake, whilst perhaps leaving some participants in the economy with a smaller slice.

We can now see why the debate over trade policy is so often contentious. Whenever a policy creates winners and losers, the stage is set for a political battle as outlined above. Nations sometimes fail to enjoy the gains from trade simply because the losers from free trade have more political influence than the winners. The losers lobby for trade restrictions, such as tariffs and import quotas.

RESTRICTIONS ON TRADE

Despite the widely acknowledged benefits that can arise from trade, the fact that there will always be winners and losers as we have seen means that arguments for restricting trade in some way are regularly promoted. We will look at three main methods of restricting trade, tariffs, quotas and non-tariff barriers and then some of the arguments for trade.

Before the tariff, the domestic price equals the world price. Consumer surplus, the area between the demand curve and the world price, is area $A + B + C + D + E + F$. Producer surplus, the area between the supply curve and the world price, is area G . Government revenue equals zero. Total surplus – the sum of consumer surplus, producer surplus and government revenue – is area $A + B + C + D + E + F + G$.

Once the government imposes a tariff, the domestic price exceeds the world price by the amount of the tariff. Consumer surplus is now area $A + B$. Producer surplus is area $C + G$. Government revenue, which is the quantity of after-tariff imports times the size of the tariff, is the area E . Thus, total surplus with the tariff is area $A + B + C + E + G$.

To determine the total welfare effects of the tariff, we add the change in consumer surplus (which is negative), the change in producer surplus (positive) and the change in government revenue (positive). We find that total surplus in the market decreases by the area $D + F$. This fall in total surplus is the *deadweight loss* of the tariff.

A tariff causes a deadweight loss simply because a tariff is a type of tax and follows the same principles as we discussed in Chapters 8 and 9. Like most taxes, it distorts incentives and pushes the allocation of scarce resources away from the optimum. In this case, we can identify two effects. First, the tariff on olive oil raises the price of olive oil that domestic producers can charge above the world price and, as a result, encourages them to increase production of olive oil (from Q_1^S to Q_2^S). Second, the tariff raises the price that domestic olive oil buyers have to pay and, therefore, encourages them to reduce consumption of olive oil (from Q_1^D to Q_2^D). Area D represents the deadweight loss from the overproduction of olive oil, and area F represents the deadweight loss from the under consumption. The total deadweight loss of the tariff is the sum of these two triangles.

The Effects of an Import Quota

The Isolandian economists next consider the effects of an **import quota** – a limit on the quantity of imports. In particular, imagine that the Isolandian government distributes a limited number of import licences. Each licence gives the licence holder the right to import 1 tonne of olive oil into Isoland from abroad. The Isolandian economists want to compare welfare under a policy of free trade and welfare with the addition of this import quota.

import quota a limit on the quantity of a good that can be produced abroad and sold domestically

The graph and table in Figure 19.13 show how an import quota affects the Isolandian market for olive oil. Because the import quota prevents Isolandians from buying as much olive oil as they want from abroad, the supply of olive oil is no longer perfectly elastic at the world price. Instead, as long as the price of olive oil in Isoland is above the world price, the licence holders import as much as they are permitted, and the total supply of olive oil in Isoland equals the domestic supply plus the quota amount. That is, the supply curve above the world price is shifted to the right by exactly the amount of the quota. (The supply curve below the world price does not shift because, in this case, importing is not profitable for the licence holders.)

The price of olive oil in Isoland adjusts to balance supply (domestic plus imported) and demand. As the figure shows, the quota causes the price of olive oil to rise above the world price. The domestic quantity demanded falls from Q_1^D to Q_2^D , and the domestic quantity supplied rises from Q_1^S to Q_2^S . Not surprisingly, the import quota reduces olive oil imports.

Now consider the gains and losses from the quota. Because the quota raises the domestic price above the world price, domestic sellers are better off, and domestic buyers are worse off. In addition, the licence holders are better off because they make a profit from buying at the world price and selling at the higher domestic price. To measure these gains and losses, we look at the changes in consumer surplus, producer surplus and licence-holder surplus.

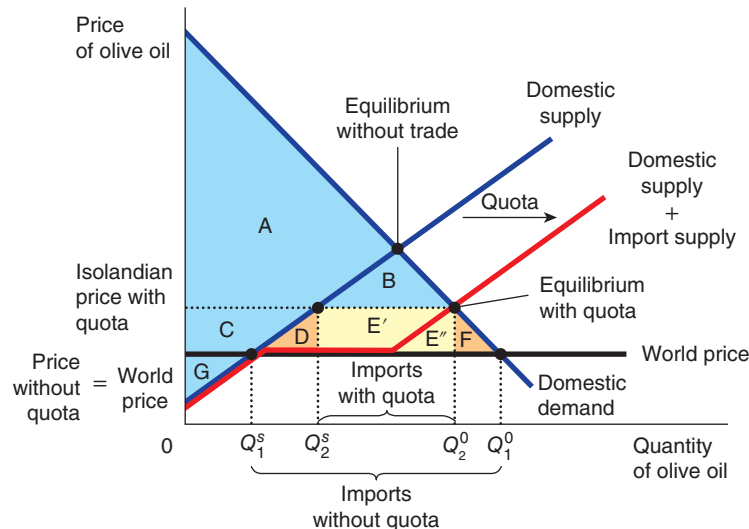
Before the government imposes the quota, the domestic price equals the world price. Consumer surplus, the area between the demand curve and the world price, is area $A + B + C + D + E' + E'' + F$. Producer surplus, the area between the supply curve and the world price, is area G . The surplus of licence

FIGURE 19.13**The Effects of an Import Quota**

An import quota, like a tariff, reduces the quantity of imports and moves a market closer to the equilibrium that would exist without trade. Total surplus falls by an amount equal to area $D + F$. These two triangles represent the deadweight loss from the quota. In addition, the import quota transfers $E' + E''$ to whoever holds the import licences.

| | Before quota | After quota | Change |
|--------------------|------------------------------------|----------------------------|---------------------------|
| Consumer surplus | $A + B + C + D + E' + E'' + F$ | $A + B$ | $-(C + D + E' + E'' + F)$ |
| Producer surplus | G | $C + G$ | $+C$ |
| Government revenue | None | $E' + E''$ | $+(E' + E'')$ |
| Total surplus | $A + B + C + D + E' + E'' + F + G$ | $A + B + C + E' + E'' + G$ | $-(D + F)$ |

The area $D + F$ shows the fall in total surplus and represents the deadweight loss of the quota.



holders equals zero because there are no licences. Total surplus, the sum of consumer, producer and licence-holder surplus, is area $A + B + C + D + E' + E'' + F + G$.

After the government imposes the import quota and issues the licences, the domestic price exceeds the world price. Domestic consumers get surplus equal to area $A + B$, and domestic producers get surplus equal to area $C + G$. The licence holders make a profit on each unit imported equal to the difference between the Isolandian price of olive oil and the world price. Their surplus equals this price differential times the quantity of imports. Thus, it equals the area of the rectangle $E' + E''$. Total surplus with the quota is the area $A + B + C + E' + E'' + G$.

To see how total welfare changes with the imposition of the quota, we add the change in consumer surplus (which is negative), the change in producer surplus (positive) and the change in licence-holder surplus (positive). We find that total surplus in the market decreases by the area $D + F$. This area represents the deadweight loss of the import quota.

This analysis should seem somewhat familiar. Indeed, if you compare the analysis of import quotas in Figure 19.13 with the analysis of tariffs in Figure 19.12, you will see that they are essentially identical. Both tariffs and import quotas raise the domestic price of the good, reduce the welfare of domestic consumers, increase the welfare of domestic producers and cause deadweight losses. There is only one difference between these two types of trade restriction: a tariff raises revenue for the government (area E in Figure 19.12), whereas an import quota creates surplus for licence holders (area $E' + E''$ in Figure 19.13).

Tariffs and import quotas can be made to look even more similar. Suppose that the government tries to capture the licence-holder surplus for itself by charging a fee for the licences. A licence to sell 1 tonne of olive oil is worth exactly the difference between the Isolandian price of olive oil and the world price, and the government can set the licence fee as high as this price differential. If the government does this, the licence fee for imports works exactly like a tariff: consumer surplus, producer surplus and government revenue are exactly the same under the two policies.

In practice, however, countries that restrict trade with import quotas rarely do so by selling the import licences. For example, in 1991 the European Union reached an agreement with Japan to 'voluntarily' limit the sale of Japanese cars in member countries of the EU. In this case, the Japanese government allocates the import licences to Japanese firms, and the surplus from these licences (area $E' + E''$) accrues to those firms. This kind of import quota is, from the standpoint of the welfare of the European Union, strictly worse than an EU tariff on imported cars. Both a tariff and an import quota raise prices, restrict trade and cause deadweight losses, but at least the tariff produces revenue for the European Union rather than for Japanese auto companies. It is perhaps not surprising, therefore, that this arrangement was terminated at the end of 1999.

Although in our analysis so far import quotas and tariffs appear to cause similar deadweight losses, a quota can potentially cause an even larger deadweight loss, depending on the mechanism used to allocate the import licences. Suppose that when Isoland imposes a quota, everyone understands that the licences will go to those who spend the most resources lobbying the Isolandian government. In this case, there is an implicit licence fee – the cost of lobbying. The revenues from this fee, however, rather than being collected by the government, are spent on lobbying expenses. The deadweight losses from this type of quota include not only the losses from overproduction (area D) and under consumption (area F) but also whatever part of the licence-holder surplus (area $E' + E''$) is wasted on the cost of lobbying.

Non-Tariff Barriers (NTBs)

Barriers to trade are sometimes not obvious but nevertheless present significant restrictions on the ability of firms to buy and sell goods from and to other countries. We will briefly outline some of the main ones.

Complex or Discriminatory Rules of Origin and Quality Conditions Countries may impose strict rules on the production of goods in its domestic market relating to technical specifications, health and safety, production standards and so on, which exporters may find difficult to meet or if they can, increase the cost of production considerably and thus make the imports less competitive against domestically produced goods. In addition, firms may be required to give precise details as to where goods come from – something which is not easy given the widespread use of many different component parts from across the globe in many cases. The country establishing the NTB may only allow goods to be imported if they adhere to strict rules of origin which exporters might not be able to meet.

Sanitary or Phyto Sanitary Conditions Firms may be required to provide details of food and plant exports which again have to adhere to strict conditions. Phyto sanitary refers to the health of plants and exporting firms may have to show that plants are free from pest and disease, and that measures have been taken to ensure the conditions of growth adhere to the standards laid down by the country setting the NTB. In the case of food exports, a country may set very high food safety standards and regulations which exporters find difficult or costly to meet.

Administrative Regulations Some countries might set up administrative procedures that have to be met prior to any goods or services coming into the country. The paperwork or 'red tape' involved can be excessive and add to the costs of the exporter which again make the firm less competitive. Some countries might require excessive or unreasonable labelling or packaging regulations or set burdensome customs entry procedures including the necessity of providing valuations of goods at the point of entry which might be challenged (given value can be a subjective issue) or certification to prove authenticity which again can lead to an increase in costs.

Currency Manipulation Some countries might implement measures which artificially influence the value of their currency, which leads to exporters finding the price they face being higher than it would otherwise be, and as a result reducing their competitiveness against domestic producers. We will see in the macroeconomics section of the book how exchange rates behave in response to government policies.

The Lessons for Trade Policy

The team of Isolandian economists can now write to the new president:

Dear Madam President,

You asked us three questions about opening up trade. These are our answers.

Question: If the government allowed Isolandians to import and export olive oil, what would happen to the price of olive oil and the quantity of olive oil sold in the domestic olive oil market?

Answer: Once trade is allowed, the Isolandian price of olive oil would be driven to equal the price prevailing around the world. If the world price is now higher than the Isolandian price, our price would rise. The higher price would reduce the amount of olive oil Isolandians consume and raise the amount of olive oil that Isolandians produce. Isoland would, therefore, become an olive oil exporter. This occurs because, in this case, Isoland would have a comparative advantage in producing olive oil.

Conversely, if the world price is now lower than the Isolandian price, our price would fall. The lower price would raise the amount of olive oil that Isolandians consume and lower the amount of olive oil that Isolandians produce. Isoland would, therefore, become an olive oil importer. This occurs because, in this case, other countries would have a comparative advantage in producing olive oil.

Question: Who would gain from free trade in olive oil and who would lose, and would the gains exceed the losses?

Answer: The answer depends on whether the price rises or falls when trade is allowed. If the price rises, producers of olive oil gain and consumers of olive oil lose. If the price falls, consumers gain and producers lose. In both cases, the gains are larger than the losses. Thus, free trade raises the total welfare of Isolandians.

Question: Should a tariff or an import quota be part of the new trade policy?

Answer: A tariff, like most taxes, has deadweight losses: the revenue raised would be smaller than the losses to the buyers and sellers. In this case, the deadweight losses occur because the tariff would move the economy closer to our current no-trade equilibrium. An import quota works much like a tariff and would cause similar deadweight losses. The best policy, from the standpoint of economic efficiency, would be to allow trade without a tariff or an import quota.

We hope you find these answers helpful as you decide on your new policy.

Your obedient servants,
Isolandian economics team

SELF TEST Draw the supply and demand curve for wool suits in the country of Autarka. When trade is allowed, the price of a suit falls from 300 to 200 grams of gold. In your diagram, what is the change in consumer surplus, the change in producer surplus and the change in total surplus? How would a tariff on suit imports alter these effects?

The Arguments for Restricting Trade

The letter from the economics team persuades the new president of Isoland to consider opening up trade in olive oil. She notes that the domestic price is now high compared to the world price. Free trade would, therefore, cause the price of olive oil to fall and hurt domestic olive oil producers. Before implementing the new policy, she asks Isolandian olive oil companies to comment on the economists' advice.

Not surprisingly, the olive oil companies are opposed to free trade in olive oil. They believe that the government should protect the domestic olive oil industry from foreign competition. Let's consider some of the arguments they might give to support their position and how the economics team would respond.

FYI



Other Benefits of International Trade

Our conclusions so far have been based on the standard analysis of international trade. As we have seen, there are winners and losers when a nation opens itself up to trade, but the gains to the winners exceed the losses of the losers. Yet the case for free trade can be made even stronger. There are several other economic benefits of trade beyond those emphasized in the standard analysis. Here, in a nutshell, are some of these other benefits:

- **Increased variety of goods.** Goods produced in different countries are not exactly the same. German beer, for instance, is not the same as American beer. Free trade gives consumers in all countries greater variety from which to choose.
- **Lower unit costs through economies of scale.** Some goods can be produced at low unit or average cost only if they are produced in large quantities. A firm in a small country cannot take full advantage of economies of scale if it can sell only in a small domestic market. Free trade gives firms access to larger world markets and allows them to realize economies of scale more fully.
- **Increased competition.** A company shielded from foreign competitors is more likely to have market power, which in turn gives it the ability to raise prices above competitive levels. This is a type of market failure. Opening up trade fosters competition and gives the invisible hand a better chance to work its magic.
- **Enhanced flow of ideas.** The transfer of technological advances around the world is often thought to be linked to international trade in the goods that embody those advances. The best way for a poor, agricultural nation to learn about the computer revolution, for instance, is to buy some computers from abroad, rather than trying to make them domestically.
- **Generates economic growth.** For poor countries, the increase in output can be a trigger to generating economic growth which may also bring an improvement in the standard of living for its citizens.

Thus, free international trade increases variety for consumers, allows firms to take advantage of economies of scale, makes markets more competitive, facilitates the spread of technology and helps generate economic growth. If the Isolandian economists thought these effects were important, their advice to their president would be even more forceful.

The Jobs Argument Opponents of free trade often argue that trade with other countries destroys domestic jobs. In our example, free trade in olive oil would cause the price of olive oil to fall, reducing the quantity of olive oil produced in Isoland and thus reducing employment in the Isolandian olive oil industry. Some Isolandian olive oil workers would lose their jobs.

Yet free trade creates jobs at the same time that it destroys them. When Isolandians buy olive oil from other countries, those countries obtain the resources to buy other goods from Isoland. Isolandian workers would move from the olive oil industry to those industries in which Isoland has a comparative advantage. Although the transition may impose hardship on some workers in the short run, it allows Isolandians as a whole to enjoy a higher standard of living.

Opponents of trade are often sceptical that trade creates jobs. They might respond that *everything* can be produced more cheaply abroad. Under free trade, they might argue, Isolandians could not be profitably employed in any industry. The gains from trade are based on comparative advantage, not absolute advantage. Even if one country is better than another country at producing everything, each country can still gain from trading with the other. Workers in each country will eventually find jobs in the industry in which that country has a comparative advantage.

The National Security Argument When an industry is threatened with competition from other countries, opponents of free trade often argue that the industry is vital for national security. Free trade would allow Isoland to become dependent on foreign countries to supply vital resources. If a war later broke out, Isoland might be unable to produce enough to defend itself and remain self-sufficient.

Economists acknowledge that protecting key industries may be appropriate when there are legitimate concerns over national security. Yet they fear that this argument may be used too quickly by producers eager to gain at consumers' expense. Certainly, it is tempting for those in an industry to exaggerate their role in national defence to obtain protection from foreign competition.

The Infant Industry Argument New industries sometimes argue for temporary trade restrictions to help them get started. After a period of protection, the argument goes, these industries will mature and be able to compete with foreign competitors. Similarly, older industries sometimes argue that they need protection to help them adjust to new conditions. In 2002, for example, when US President George Bush imposed steep tariffs on the import of steel from the European Union, he argued that the industry needed protection in order to be able to afford to pay the pensions and health care costs of its retired workers and while it was going through a period of adjustment in terms of making its production more efficient in order to be able to cope with intense foreign competition.

Economists are often sceptical about such claims. The primary reason is that the infant industry argument is difficult to implement in practice. To apply protection successfully, the government would need to decide which industries will eventually be profitable and decide whether the benefits of establishing these industries exceed the costs to consumers of protection. Yet 'picking winners' is extraordinarily difficult. It is made even more difficult by the political process, which often awards protection to those industries that are politically powerful. And once a powerful industry is protected from foreign competition, the 'temporary' policy is hard to remove.

In addition, many economists are sceptical about the infant industry argument even in principle. Suppose, for instance, that the Isolandian olive oil industry is young and unable to compete profitably against foreign rivals. Yet there is reason to believe that the industry can be profitable in the long run. In this case, the owners of the firms should be willing to incur temporary losses to obtain the eventual profits. Protection is not necessary for an industry to grow. Firms in various industries – such as many internet firms today – incur temporary losses in the hope of growing and becoming profitable in the future. And many of them succeed, even without protection from foreign competition.

The Unfair Competition Argument A common argument is that free trade is desirable only if all countries play by the same rules. If firms in different countries are subject to different laws and regulations, then it is unfair (the argument goes) to expect the firms to compete in the international marketplace. For instance, suppose that the government of Neighbourland subsidizes its olive oil industry by giving olive oil companies large tax breaks. The Isolandian olive oil industry might argue that it should be protected from this foreign competition because Neighbourland is not competing fairly.

Would it, in fact, hurt Isoland to buy olive oil from another country at a subsidized price? Certainly, Isolandian olive oil producers would suffer, but Isolandian olive oil consumers would benefit from the low price. Moreover, the case for free trade is no different: the gains of the consumers from buying at the low price would exceed the losses of the producers. Neighbourland's subsidy to its olive oil industry may be a bad policy, but it is the taxpayers of Neighbourland who bear the burden. Isoland can benefit from the opportunity to buy olive oil at a subsidized price.

The Protection as a Bargaining Chip Argument Another argument for trade restrictions concerns the strategy of bargaining. Many policymakers claim to support free trade but, at the same time, argue that trade restrictions can be useful when we bargain with our trading partners. They claim that the threat of a trade restriction can help remove a trade restriction already imposed by a foreign government. For example, Isoland might threaten to impose a tariff on olive oil unless Neighbourland removes its tariff on wheat. If Neighbourland responds to this threat by removing its tariff, the result can be freer trade.

The problem with this bargaining strategy is that the threat may not work. If it doesn't work, the country has a difficult choice. It can carry out its threat and implement the trade restriction, which would reduce its own economic welfare. Or it can back down from its threat, which would cause it to lose prestige in international affairs. Faced with this choice, the country would probably wish that it had never made the threat in the first place.

CONCLUSION

The principle of comparative advantage shows that trade can make everyone better off. You should now understand more fully the benefits of living in an interdependent economy. Despite the benefits of trade, there are fierce debates about trade policy. Economists and the general public often disagree about free trade. In particular, economists are usually opposed to trade restrictions while the general public often wants to see restrictions on trade in order to 'protect' the domestic economy from 'cut-throat' foreign competition.

To understand better the economists' view of trade, let's continue our parable. Suppose that the country of Isoland ignores the advice of its economics team and decides not to allow free trade in olive oil. The country remains in the equilibrium without international trade.

Then, one day, some Isolandian inventor discovers a new way to make olive oil at very low cost. The process is quite mysterious, however, and the inventor insists on keeping it a secret. What is odd is that the inventor doesn't need any workers or fields to make olive oil. The only input he requires is wheat.

The inventor is hailed as a genius. Because olive oil is used in so many products, the invention lowers the cost of many goods and allows all Isolandians to enjoy a higher standard of living. Workers who had previously produced olive oil do suffer when their factories close, but eventually they find work in other industries. Some become farmers and grow the wheat that the inventor turns into olive oil. Others enter new industries that emerge as a result of higher Isolandian living standards. Everyone understands that the displacement of these workers is an inevitable part of progress.

After several years, a newspaper reporter decides to investigate this mysterious new olive oil process. She sneaks into the inventor's factory and learns that the inventor is a fraud. The inventor has not been making olive oil at all. Instead, he has been smuggling wheat abroad in exchange for olive oil from other countries. The only thing that the inventor had discovered was the gains from international trade.

When the truth is revealed, the government shuts down the inventor's operation. The price of olive oil rises and workers return to jobs in olive oil factories. Living standards in Isoland fall back to their former levels. The inventor is jailed and held up to public ridicule. After all, he was no inventor. He was just an economist.

IN THE NEWS

Comparative Advantage

Some economic theories, like the principle of comparative advantage, seem to have been passed down through the generations as received wisdom. Yet, economics is a dynamic subject and theories, however well respected they might be, should never be above critique. This article presents a view of comparative advantage in a different light.

Time to Review Economics Curriculum

Economics cannot be separated from everyday activities of man. Almost everything in life has an economic content. In the field of economics, no aspect of human behaviour is off limit in the search

for the key that unlocks economic prosperity. From the days of Adam Smith, economists have continued to be very influential in the world. Their opinions and analyses still wield enormous impact on people's lives. The importance of this subject area is such that virtually all presidents have economic advisers.

No other human challenge/discipline shares the same podium with economics in terms of policy priority in any government.

However, in the recent days, economics is in danger of losing its influence. One of the major revelations from the global financial crisis is that macroeconomics did not keep

(Continued)

pace with the revolution in finance, to the great detriment of contemporary economics. Paul Krugman, the 2008 winner of the Nobel Prize in economics, described advances in the field of macroeconomics in the past 30 years as 'spectacularly useless at best and positively harmful at worst'. It is therefore not surprising that in January 2011, the theme of the annual general meeting of the American Economic Association centred on the failings of the discipline as highlighted by the recent global financial crisis.

One of the major reasons economics is losing relevance is that the simplifying assumptions upon which mainstream economic theories and principles are based are no longer useful in explaining the ever-changing real world developments. Western economists, in pushing for free market policies, often ignore ways in which the real world departs from the idealized one of neoclassical economic theory that currently dominates the study and the teaching of economics. The West's insistence that economic policy models must not deviate too widely from the free market models greatly reduces the prospect of other models and propositions which have more solid empirical evidence supporting them. The belief that whatever free market models leave out does not matter opens the policy arena to dubious views that ridicule the profession. Unfortunately, it is these faulty models/principles that still hold sway in our schools, more as a result of elegance than relevance.

A case in point is the principle of comparative advantage. This principle asserts that a country 'should and under competitive conditions specialize in the export of the products that it can produce at the lowest relative cost'. This principle, according to Swedish economists,

Eli Heckscher and Bertil Ohlin, advocates that countries should specialize in line with their factor endowments. It implies that developing countries endowed with abundant supplies of unskilled labour should specialize in the export of primary products that will exploit the pool of their unskilled labour. It further argues that by 'trading these primary products for the manufactured goods that developed countries were theoretically best suited to produce, developing countries could realize the enormous potential benefits from free trade with the richer nations of the world'. This principle, which first appeared in a book called *On The Principles of Political Economy and Taxation* by David Ricardo in 1817, is still being taught in our secondary and tertiary institutions today, not rightly as a topic in the 'History of Economic Thoughts' lectures, but wrongly as a core topic in the 'International Trade and Development' lecture series.

History has shown that almost all nations that moved from poverty to prosperity went against the principles of comparative advantage and those that imbibed the principle got trapped in the vicious cycle of poverty. For over 300 years, Africa has been exporting primary commodities, yet the region is the poorest continent in terms of all human development indices. The level of underdevelopment in the continent of Africa is in sharp contrast with the level of economic transformation witnessed in China, Taiwan, South Korea, India and Singapore in the past 60 years. These countries representing the world's most successful economies in the past four decades transformed their nations by implementing policies that were at variance with the concept of comparative advantage. It is clear that these countries' resource

endowments and cost structure do not predispose them to specialize in any particular industry where they currently hold the ace. There is no factor-endowment that would have made India, for example, have a comparative advantage in software technology. The truth is that when a country develops new products or new methods, it is creating a comparative advantage that it did not have before. Economies develop in a particular way because 'someone somewhere makes a conscious decision to invest, acquire skills and gain knowledge in particular areas'. President Park Chung Hee chose to spur South Korea's industrialization through large businesses. Through his government's support and finance, great brands like Samsung, Hyundai, and LG were nurtured. These conglomerates helped South Korea become one of the most industrialized nations in the world. Lee Kuan Yew, the first prime minister of independent Singapore, is credited with transforming the small island nation from a developing economy to one of the advanced economies of the world. Today, Singapore, with only 3.8 million inhabitants, reportedly exports more manufactured products than all sub-Saharan African countries together. Other examples abound.

It is therefore absurd that in an era when statistics have shown that world top ten exporters of manufactured/processed goods account for about 50 per cent of world's exports, while 50 least developed countries account for less than 0.5 per cent of global non-oil export (a level virtually unchanged over the past 15 years), we are still teaching our students that there is an economic gain in specializing on the production of primary products that are either finite in supply or dependent on weather conditions. Rather than

teach that specialization should be based on comparative advantage, we should teach that specialization should be based on activities where benefits are likely to be highest. Our economics students must be made to understand that what a country produces matters. No nation has joined the league of advanced countries by producing and exporting only primary commodities. As it is often said, it is not what people do not know that is the problem; the greatest danger lies in what they believe to be true that is not true.

It is really difficult to understand why alternative principles like the Prebisch–Singer thesis are relegated to the background in our economics curriculum. Economists Raul Prebisch and Hans Singer independently developed the hypothesis that people tend to spend more on durable sophisticated manufactured goods as income goes up but less on primary products even when prices drop. They therefore assert that trade with producers of manufactured goods will result in declining

terms of trade for countries producing primary goods, and every student of international trade knows that countries have incentives to improve their terms of trade. Several other economists have proved from pragmatic, theoretical and empirical grounds that the premature free trade imposed on developing nations according to the doctrine of comparative advantage is a major cause of the persistent poverty in the developing world. Yet it is the dubious ‘principle of comparative advantage’ that we recite in our classrooms.

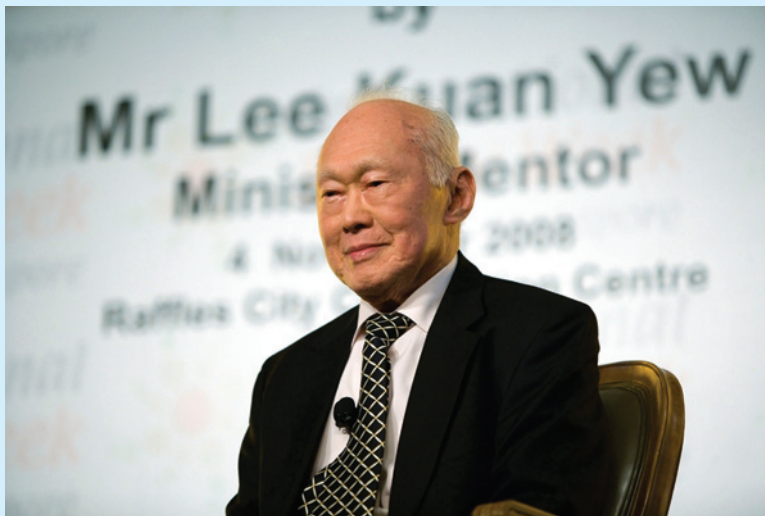
Sadly, this primitive principle has endured long enough in our academics that it seems to have entrenched itself into our policymaking circle at the expense of its more empirical sound descendants. Otherwise, how can we explain that, 52 years after independence, Nigeria still exports crude oil to import refined petroleum products? As John M. Keynes pointed out, ‘The economists who most influence practical men of action are the defunct ones whose scribbling have had time to percolate from the

seminar room to wider conversation.’ It is now time to expunge them from our curriculum. This is urgent because an economist’s worldview is naturally influenced by the theories and principles he is taught. Since ideological factors were robust predictors of managers’ reactions to scenarios, an economist trained with economic principles and theories that lack empirical and historical validity is a national risk.

Questions

- 1 To what extent would you agree with the view that theories like comparative advantage are based on too simplistic a set of assumptions?
- 2 Explain how countries in Africa, which have an abundance of labour as a natural resource, can benefit from trade and allow countries to grow by specializing in the export of primary products.
- 3 Do you think the author is making his own assumptions that countries like South Korea and Singapore have somehow bypassed their natural comparative advantage and moved to compete in other areas to achieve growth? (Hint: does comparative advantage always assume that development must start with the export of primary goods?)
- 4 What other factors might have caused countries in Africa to have not exploited their comparative advantage and as such have remained relatively poor in comparison to some of the more rapidly developed nations quoted in the article?
- 5 To what extent would you agree that an alternative theory to comparative advantage ought to be included in economics curricula?

Lee Kuan Yew, the first prime minister of independent Singapore, is credited with transforming the small island nation from a developing economy to one of the advanced economies of the world.



SUMMARY

- Production possibilities frontiers provide a model to show potential output of goods and services in an economy and the opportunity cost ratios of diverting resources to different uses.
- The PPF can shift outwards if countries find ways of improving their factor productivity or exploit factor endowments more effectively.
- The shape and position of a PPF is dependent on the productivity of factor inputs and degree of specialization involved in the country in different industries.
- Each person consumes goods and services produced by many other people both in their country and around the world. Interdependence and trade are desirable because they allow everyone to enjoy a greater quantity and variety of goods and services.
- There are two ways to compare the ability of two people in producing a good. The person who can produce the good with the smaller quantity of inputs is said to have an *absolute advantage* in producing the good. The person who has the smaller opportunity cost of producing the good is said to have a *comparative advantage*. The gains from trade are based on comparative advantage, not absolute advantage.
- Trade makes everyone better off because it allows people to specialize in those activities in which they have a comparative advantage.
- The principle of comparative advantage applies to countries as well as to people. Economists use the principle of comparative advantage to advocate free trade among countries.
- The effects of free trade can be determined by comparing the domestic price without trade to the world price. A low domestic price indicates that the country has a comparative advantage in producing the good and that the country will become an exporter. A high domestic price indicates that the rest of the world has a comparative advantage in producing the good and that the country will become an importer.
- When a country allows trade and becomes an exporter of a good, producers of the good are better off, and consumers of the good are worse off. When a country allows trade and becomes an importer of a good, consumers are better off, and producers are worse off. In both cases, the gains from trade exceed the losses.
- A tariff – a tax on imports – moves a market closer to the equilibrium that would exist without trade and, therefore, reduces the gains from trade. Although domestic producers are better off and the government raises revenue, the losses to consumers exceed these gains.
- An import quota – a limit on imports – has effects that are similar to those of a tariff. Under a quota, however, the holders of the import licences receive the revenue that the government would collect with a tariff.
- There are various arguments for restricting trade: protecting jobs, defending national security, helping infant industries, preventing unfair competition and responding to foreign trade restrictions. Although some of these arguments have some merit in some cases, economists believe that free trade is usually the better policy.

QUESTIONS FOR REVIEW

- 1 Draw a PPF for a country producing only computers and wheat. What determines the shape and position of the PPF you have drawn? Use your diagram to show the opportunity cost of different output combinations.
 - 2 What does the domestic price that prevails without international trade tell us about a nation's comparative advantage?
 - 3 Explain how absolute advantage and comparative advantage differ.
 - 4 Can two countries gain from trade if the opportunity cost ratios relating to production of goods they can both produce is the same? Explain.
 - 5 When does a country become an exporter of a good? An importer?
 - 6 Thousands of tourists visit Europe cities every year. Is tourism an import or an export to a country? Explain.
 - 7 Draw the supply and demand diagram for an importing country. What is consumer surplus and producer surplus before trade is allowed? What is consumer surplus and producer surplus with free trade? What is the change in total surplus?
 - 8 Describe what a tariff is, and describe its economic effects.
 - 9 What is an import quota? Compare its economic effects with those of a tariff.
 - 10 List five arguments often given to support trade restrictions. How do economists respond to these arguments?
-

PROBLEMS AND APPLICATIONS

- 1 Draw the PPF for a country which produces just oranges and cars. Assume that the country is better at producing oranges than cars.
 - a. What can you say about the opportunity cost of the country diverting more resources into producing cars?
 - b. The country discovers new resources which will mean that its ability to produce cars is significantly improved. What happens to the shape of the PPF as a result of this discovery?
 - c. What can you say about the opportunity cost facing the country after this new discovery?
- 2 Look at the table below which shows the production possibilities for two products, pens and pencils.

| Point | Pens | Pencils |
|-------|------|---------|
| A | 10 | 0 |
| B | 5 | 10 |
| C | 2 | 25 |

- a. Calculate the opportunity cost of producing 1 additional pen between points A and B.
 - b. Calculate the opportunity cost of producing 1 additional pencil between points A and B.
 - c. Calculate the opportunity cost of producing 1 additional pen between points B and C. What conclusions can you come to about this opportunity cost compared to the calculation you made in a. above?
 - d. Calculate the opportunity cost of producing 1 additional pencil between points B and C. What conclusions can you come to about this opportunity cost compared to the calculation you made in b. above?
- 3 Swedish and Danish workers can each produce 4 capital goods a year. A Swedish worker can produce 10 tonnes of grain a year, whereas a Danish worker can produce 5 tonnes of grain a year. To keep things simple, assume that each country has 100 million workers.
 - a. For this situation, construct a table analogous to Table 19.1.
 - b. Graph the production possibilities frontier of the Swedish and Danish economies.
 - c. For Sweden, what is the opportunity cost of a car? Of grain? For Denmark, what is the opportunity cost of a car? Of grain? Put this information in a table analogous to Table 19.3.
 - d. Which country has an absolute advantage in producing capital goods? In producing grain?
 - e. Which country has a comparative advantage in producing capital goods? In producing grain?

- f. Without trade, half of each country's workers produce capital goods and half produce grain. What quantities of capital goods and grain does each country produce?
 - g. Starting from a position without trade, give an example in which trade makes each country better off.
- 4 Morgan and Oliver share a flat. They spend most of their time studying (of course), but they leave some time for their favourite activities: cooking pizza and making homebrew beer. Oliver takes 4 hours to produce 1 barrel of homebrew and 2 hours to make a pizza. Morgan takes 6 hours to brew 1 barrel of beer and 4 hours to make a pizza.
 - a. What is each flatmate's opportunity cost of making a pizza? Who has the absolute advantage in making pizza? Who has the comparative advantage in making pizza?
 - b. If Oliver and Morgan trade foods with each other, who will trade away pizza in exchange for homebrew?
 - c. The price of pizza can be expressed in terms of barrels of homebrew. What is the highest price at which pizza can be traded that would make both flatmates better off? What is the lowest price? Explain.
 - 5 The UK and Poland both produce cakes and coats. Suppose that a British worker can produce 50 cakes per hour or 1 coat per hour. Suppose that a Polish worker can produce 40 cakes per hour or 2 coats per hour.
 - a. Which country has the absolute advantage in the production of each good? Which country has the comparative advantage?
 - b. If the UK and Poland decide to trade, which commodity will Poland trade to the UK? Explain.
 - c. If a Polish worker could produce only 1 coat per hour, would Poland still gain from trade? Would the UK still gain from trade? Explain.
 - 6 France represents a small part of the world apple market.
 - a. Draw a diagram depicting the equilibrium in the French apple market without international trade. Identify the equilibrium price, equilibrium quantity, consumer surplus and producer surplus.
 - b. Suppose that the world apple price is below the French price before trade, and that the French apple market is now opened to trade. Identify the new equilibrium price, quantity consumed, quantity produced domestically and quantity imported. Also show the change in the surplus of domestic consumers and producers. Has domestic total surplus increased or decreased?

- 7** The world price of wine is below the price that would prevail in France in the absence of trade.
- Assuming that French imports of wine are a small part of total world wine production, draw a graph for the French market for wine under free trade. Identify consumer surplus, producer surplus and total surplus in an appropriate table.
 - Now suppose that an outbreak of phyloxera (a sap sucking insect which damages grape vines) in California and South America destroys much of the grape harvest there. What effect does this shock have on the world price of wine? Using your graph and table from part (a), show the effect on consumer surplus, producer surplus and total surplus in France. Who are the winners and losers? Is France better or worse off?
- 8** The world price of cotton is below the no-trade price in country A and above the no-trade price in country B. Using supply and demand diagrams and welfare tables such as those in the chapter, show the gains from trade in each country. Compare your results for the two countries.
- 9** Suppose that European Union countries impose a common tariff on imported cars to protect the European car industry from foreign competition. Assuming that Europe is a price taker in the world car market, show on a diagram: the change in the quantity of imports, the loss to European consumers, the gain to European car manufacturers, government revenue and the deadweight loss associated with the tariff. The loss to consumers can be decomposed into three pieces: a transfer to domestic producers, a transfer to the government and a deadweight loss. Use your diagram to identify these three pieces.
- 10** When the government of Tradeland decides to impose an import quota on foreign cars, three proposals are suggested: sell the import licences in an auction. Distribute the licences randomly in a lottery. Let people wait in line and distribute the licences on a first-come, first-served basis. Compare the effects of these policies. Which policy do you think has the largest deadweight losses? Which policy has the smallest deadweight losses? Why? (Hint: the government's other ways of raising tax revenue themselves all cause deadweight losses.)

PART 10

THE DATA OF

MACROECONOMICS

20 MEASURING A NATION'S INCOME

When you finish university and start looking for a full-time job, your experience will, to a large extent, be shaped by prevailing economic conditions. In some years, firms throughout the economy are expanding their production of goods and services, employment is rising, and jobs are relatively easy to find. In other years, firms are cutting back on production, employment is declining, and finding a good job takes a long time. Not surprisingly, any university graduate would rather enter the labour force in a year of economic expansion than in a year of economic contraction.

Because the condition of the overall economy profoundly affects all of us, changes in economic conditions are widely reported by the media. Indeed, it is hard to pick up a newspaper without seeing some newly reported statistic about the economy. The statistic might measure the total income of everyone in the economy called gross domestic product (GDP), the rate at which average prices are rising (inflation), the percentage of the labour force that is out of work (unemployment), total spending in shops (retail sales), or the imbalance of trade between the domestic economy and the rest of the world (the trade deficit). All these statistics are *macroeconomic*. Rather than telling us about a particular household or firm, they tell us something about the entire economy.

Economics is divided into two branches: microeconomics and macroeconomics. Microeconomics is the study of how individual households and firms make decisions and how they interact with one another in markets. Macroeconomics is the study of the economy as a whole. The goal of macroeconomics is to explain the economic changes that affect many households, firms and markets simultaneously. Macroeconomists address diverse questions: why is average income high in some countries while it is low in others? Why do prices rise rapidly in some periods of time while they are more stable in other periods? Why do production and employment expand in some years and contract in others? What, if anything, can the government do to promote rapid growth in incomes, low inflation and stable employment? These questions are all macroeconomic in nature because they concern the workings of the entire economy.

Because the economy as a whole is just a collection of many households and many firms interacting in many markets, microeconomics and macroeconomics are closely linked. The basic tools of supply and demand, for instance, are as central to macroeconomic analysis as they are to microeconomic analysis. Yet studying the economy in its entirety raises some new and intriguing challenges.

In this chapter and the next one, we discuss some of the data that economists and policymakers use to monitor the performance of the overall economy. These data reflect the economic changes that

macroeconomists try to explain. This chapter considers *gross domestic product*, or simply GDP, which measures the total income of a nation. GDP is the most closely watched economic statistic because it is thought to be the best single (although not perfect) measure of a society's economic well-being.

THE ECONOMY'S INCOME AND EXPENDITURE

If you were to judge how a person is doing economically, you might first look at his or her income. A person with a high income can more easily afford life's necessities and luxuries. It is no surprise that people with higher incomes enjoy higher standards of living – better housing, better health care, fancier cars, more opulent vacations and so on.

The same logic applies to a nation's overall economy. When judging whether the economy is doing well or poorly, it is natural to look at the total income that everyone in the economy is earning. This is the task of gross domestic product (GDP).

GDP measures two things at once: the total income of everyone in the economy and the total expenditure on the economy's output of goods and services. The reason that GDP can perform the trick of measuring both total income and total expenditure is that these two things are really the same. For an economy as a whole, income must equal expenditure.

Why is this true? An economy's income is the same as its expenditure because every transaction has two parties: a buyer and a seller. Every euro of spending by some buyer is a euro of income for some seller. Suppose, for instance, that Michael pays Astrid €20 to clean the windows at his house. In this case, Astrid is a seller of a service and Michael is a buyer. Astrid earns an income of €20 and Michael spends €20. Thus, the transaction contributes equally to the economy's income and to its expenditure. GDP, whether measured as total income or total expenditure, rises by €20.

Another way to see the equality of income and expenditure is with the circular-flow diagram in Figure 20.1. We presented a simple circular-flow diagram in Chapter 2 but this one includes some additional information. You may remember that this is a model which describes all the transactions between households and firms in a simple economy. In this economy, households buy goods and services from firms; these expenditures flow through the markets for goods and services. The firms in turn use the money they receive from sales to pay workers' wages, landowners' rent and firm owners' profit; this income flows through the markets for the factors of production. In this economy, money flows from households to firms and then back to households. The additional information in this circular flow diagram covers what are termed leakages and injections from and into the circular flow.

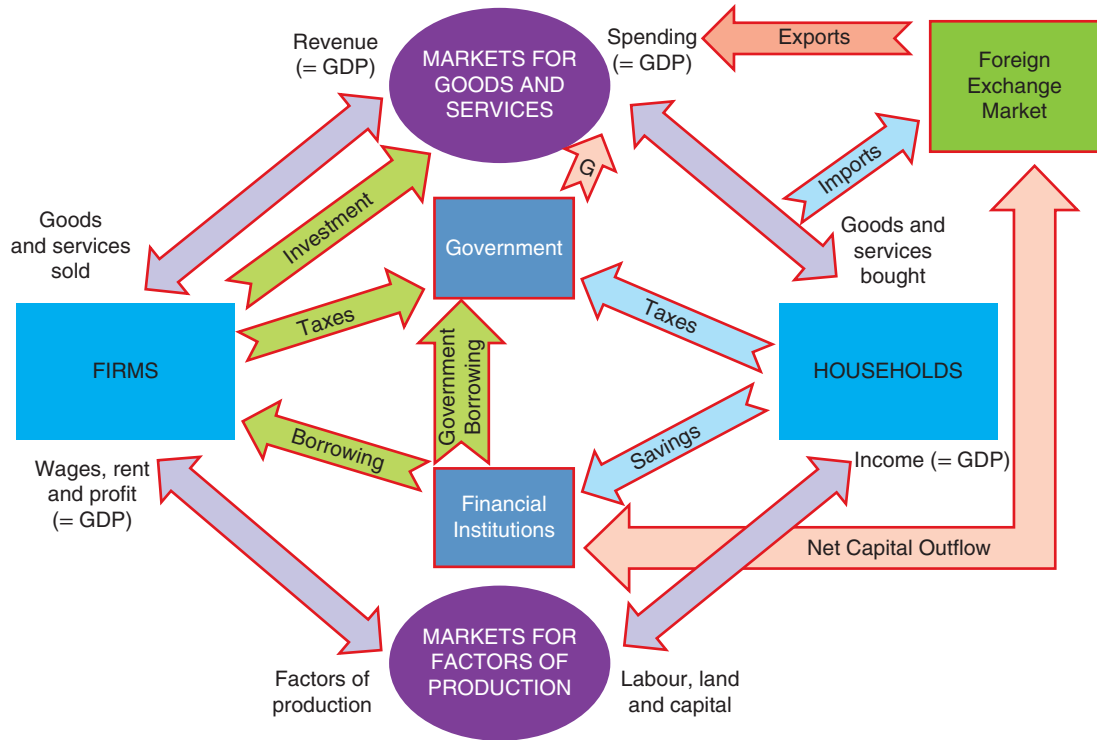
When households receive income some of this income is subject to tax. Some households will not spend all the income they receive, some of it is saved and the funds find their way into financial institutions in the form of pension saving, insurance and assurance, and saving in bank accounts. Some of the income spent on goods and services by households leaves the economy in the form of spending on imports. Similarly, some of the income received by firms is paid to the government in business taxes. Taxes (T), saving (S) and spending on imports (M) are leakages to the circular flow.

Governments use the funds received in tax revenue and borrowing from financial institutions to spend on government services and investment on infrastructure, education, health, defence and so on. Firms also make use of the funds provided by savers in financial institutions by borrowing to finance investment spending on new plant, equipment and expansion. This investment finds its way back into the market for goods and services. Firms will also sell some of the goods and services they produce abroad and so the revenue from exports flows back into the system. Financial institutions also lend money overseas and firms overseas invest money into the country which is recorded as net capital outflow.

The injections to the circular flow, therefore, are government spending (G), investment spending (I) and revenue from exports (X). We can compute GDP for this economy by adding the sum of consumer spending, government spending, investment spending and the difference between the amount paid out in imports and that received in exports (net exports). Regardless of whether a household, government or firm buys a good or service, the transaction has a buyer and seller. Thus, for the economy as a whole, expenditure and income are always the same.

FIGURE 20.1**The Circular-Flow Diagram**

Households buy goods and services from firms, and firms use their revenue from sales to pay wages to workers, rent to landowners and profit to firm owners. GDP equals the total amount spent by households in the market for goods and services. It also equals the total wages, rent and profit paid by firms in the markets for the factors of production. Leakages to the circular flow include taxes, savings and spending on imports whereas injections to the circular flow include government spending, investment and revenue generated from exports.



SELF TEST What two things do gross domestic product measure? How can it measure two things at once?

THE MEASUREMENT OF GROSS DOMESTIC PRODUCT

Now that we have discussed the meaning of gross domestic product in general terms, let's be more precise about how this statistic is measured. Here is a definition of GDP:

- **Gross domestic product (GDP)** is the market value of all final goods and services produced within a country in a given period of time.

gross domestic product (GDP) the market value of all final goods and services produced within a country in a given period of time

This definition might seem simple enough. But, in fact, many subtle issues arise when computing an economy's GDP. Let's therefore consider each phrase in this definition with some care.

'GDP Is the Market Value ...'

You have probably heard the adage, 'You can't compare apples and oranges.' Yet GDP does exactly that. GDP adds together many different kinds of products into a single measure of the value of economic activity. To do this, it uses market prices. Because market prices measure the amount people are willing to pay for different goods, they reflect the value of those goods. If the price of an apple is twice the price of an orange, then an apple contributes twice as much to GDP as does an orange.

'... Of All ...'

GDP tries to be comprehensive. It includes all items produced in the economy and sold legally in markets. GDP measures the market value of not just apples and oranges, but also pears and grapefruit, books and movies, haircuts and health care, and so on.

GDP also includes the market value of the housing services provided by the economy's stock of housing. For rental housing, this value is easy to calculate – the rent equals both the tenant's expenditure and the landlord's income. Yet many people own the place where they live and, therefore, do not pay rent. The government includes this owner-occupied housing in GDP by estimating its rental value. That is, GDP is based on the assumption that the owner, in effect, pays rent to himself, so the rent is included both in his expenditure and in his income.

There are some products, however, that GDP excludes because measuring them is so difficult. GDP excludes most items produced and sold illicitly, such as illegal drugs. It also excludes most items that are produced and consumed at home and, therefore, never enter the marketplace. Vegetables you buy at the greengrocer's shop or the supermarket are part of GDP; vegetables you grow in your garden are not.

These exclusions from GDP can at times lead to paradoxical results. The value of work carried out by housewives and husbands is not included in GDP calculations and neither is the value of child care work carried out by grandparents, for example. In addition, when Eva pays Didi to wash her car, that transaction is part of GDP. If Eva were to marry Didi, the situation would change. Even though Didi may continue to wash Eva's car, the value of the car wash is now left out of GDP because Didi's service is no longer sold in a market. Thus, when Eva and Didi marry, GDP falls.

'... Final ...'

When a paper company sells paper to a greetings card company, the paper is called an *intermediate good*, and the card is called a *final good*. GDP includes only the value of final goods. The reason is that the value of intermediate goods is already included in the prices of the final goods. Adding the market value of the paper to the market value of the card would be double counting. That is, it would (incorrectly) count the paper twice.

An important exception to this principle arises when an intermediate good is produced and, rather than being used, is added to a firm's inventory of goods to be used or sold at a later date. In this case, the intermediate good is taken to be 'final' for the moment, and its value as inventory investment is added to GDP. When the inventory of the intermediate good is later used or sold, the firm's inventory investment is negative, and GDP for the later period is reduced accordingly.

'... Goods and Services ...'

GDP includes both tangible goods (food, clothing, cars) and intangible services (haircuts, house cleaning, doctor's visits). When you (legally) download an album by your favourite band, you are buying a good, and the purchase price is part of GDP. When you pay to hear a concert by the same band, you are buying a service, and the ticket price is also part of GDP.

'... Produced ...'

GDP includes goods and services currently produced. It does not include transactions involving items produced in the past. When Aston Martin produces and sells a new car, the value of the car is included in the GDP of the country in which Aston Martin operates. When one person sells a used car to another person, the value of the used car is not included in GDP.

'... Within a Country ...'

GDP measures the value of production within the geographic confines of a country. When an Australian citizen works temporarily in the United Kingdom, his production is part of UK GDP. When a UK citizen owns a factory in Bulgaria, the production at his factory is not part of UK GDP (it's part of Bulgaria's GDP). Thus, items are included in a nation's GDP if they are produced domestically, regardless of the nationality of the producer.

'... In a Given Period of Time'

GDP measures the value of production that takes place within a specific interval of time. Usually that interval is a year or a quarter (three months). GDP measures the economy's flow of income and expenditure during that interval.

When the government reports the GDP for a quarter, it usually presents GDP 'at an annual rate'. This means that the figure reported for quarterly GDP is the amount of income and expenditure during the quarter multiplied by 4. The government uses this convention so that quarterly and annual figures on GDP can be compared more easily.

In addition, when the government reports quarterly GDP, it presents the data after they have been modified by a statistical procedure called *seasonal adjustment*. The unadjusted data show clearly that the economy produces more goods and services during some times of year than during others. (As you might guess, December's holiday shopping season is a high point in many countries whilst the period before Ramadan is a high point for many Muslim countries.) When monitoring the condition of the economy, economists and policymakers often want to look beyond these regular seasonal changes. Therefore, government statisticians adjust the quarterly data to take out the seasonal cycle. The GDP data reported in the news are always seasonally adjusted.

Now let's repeat the definition of GDP:

- Gross domestic product (GDP) is the market value of all final goods and services produced within a country in a given period of time.

It should be apparent that GDP is a sophisticated measure of the value of economic activity. In advanced courses in macroeconomics you will learn more about the subtleties that arise in its calculation. But even now you can see that each phrase in this definition is packed with meaning.

SELF TEST Which contributes more to GDP – the production of €1 of tin or the production of €1 of gold? Why?

THE COMPONENTS OF GDP

Spending in the economy takes many forms. At any moment, the Müller family may be having lunch in a Munich restaurant; Honda may be building a car factory on the banks of the Rhine and the German Army may be procuring weapons from German arms manufacturers. German GDP includes all of these various forms of spending on domestically produced goods and services. Similarly, each country in Europe will monitor the forms of spending and income to arrive at the GDP for that country.

FYI



Other Measures of Income

When the Office for National Statistics (ONS) or Eurostat computes the GDP every three months for the UK and the EU respectively, they also compute various other measures of income to arrive at a more complete picture of what's happening in the economy. These other measures differ from GDP by excluding or including certain categories of income. What follows is a brief description of five of these income measures, ordered from largest to smallest, based on the UK.

- **Gross national product (GNP)** is the total income earned by a nation's permanent residents (called nationals). It differs from GDP by including income that domestic citizens earn abroad and excluding income that foreigners earn in the domestic country, as we saw above. For most countries, domestic residents are responsible for most domestic production, so GDP and GNP are quite close.
- **Net national product (NNP)** is the total income of a nation's residents

(GNP) minus losses from depreciation. *Depreciation* is the wear and tear on the economy's stock of equipment and structures, such as lorries rusting and computers becoming obsolete.

- **National income** is the total income earned by a nation's residents in the production of goods and services. It differs from net national product by excluding indirect business taxes (such as sales taxes) and including business subsidies. NNP and national income also differ because of a 'statistical discrepancy' that arises from problems in data collection.
- **Personal income** is the income that households and non-corporate businesses receive. Unlike national income, it excludes *retained earnings*, which is income that corporations have earned but have not paid out to their owners. It also subtracts corporate income taxes and contributions for social insurance. In addition, personal income

includes the interest income that households receive from their holdings of government debt and the income that households receive from government transfer programmes, such as welfare and social security payments.

- **Disposable personal income** is the income that households and non-corporate businesses have left after satisfying all their obligations to the government. It equals personal income minus personal taxes and certain non-tax payments (such as parking tickets).

Although the various measures of income differ in detail, they almost always tell the same story about economic conditions. When GDP is growing rapidly, these other measures of income are usually growing rapidly. And when GDP is falling, these other measures are usually falling as well. For monitoring fluctuations in the overall economy, it does not matter much which measure of income we use.

To understand how the economy is using its scarce resources, economists are often interested in studying the composition of GDP among various types of spending as we noted in our look at the circular flow of income. To do this, GDP (which we denote as Y) is divided into four components: consumption (C), investment (I), government purchases (G) and net exports (NX) – the difference of the value of exports minus the value of (expenditure on) imports ($X - M$):

$$Y \equiv C + I + G + NX$$

This equation is an *identity* – an equation that must be true by the way the variables in the equation are defined. (That's why we used the three-bar, 'identically equals' symbol, ' \equiv ', although for the most part we'll follow normal practice in dealing with identities and use the usual equals sign, ' $=$ '.) In this case, because each pound or euro of expenditure included in GDP is placed into one of the four components of GDP, the total of the four components must be equal to GDP. Let's look at each of these four components more closely.

Consumption

Consumption is spending by households on goods and services. 'Goods' include household spending on durable goods, such as cars and appliances like washing machines and fridges, and non-durable goods, such as food and clothing.

consumption spending by households on goods and services, with the exception of purchases of new housing

'Services' include such intangible items as haircuts, entertainment and medical care. Household spending on education is also included in consumption of services (although one might argue that it would fit better in the next component).

Investment

Investment is the purchase of goods that will be used in the future to produce more goods and services. It is the sum of purchases of capital equipment, inventories and structures. Investment in structures includes expenditure on new housing. By convention, the purchase of a new house is the one form of household spending categorized as investment rather than consumption.

investment spending on capital equipment, inventories and structures, including household purchases of new housing

As mentioned earlier in this chapter, the treatment of inventory accumulation is noteworthy. When Sony produces a phone and, instead of selling it, adds it to its inventory, Sony is assumed to have 'purchased' the phone for itself. That is, the national income accountants treat the phone as part of Sony's investment spending. (If Sony later sells the phone out of inventory, Sony's inventory investment will then be negative, offsetting the positive expenditure of the buyer.) Inventories are treated this way because one aim of GDP is to measure the value of the economy's production, and goods added to inventory are part of that period's production.

Government Purchases

Government purchases include spending on goods and services by local and national governments. It includes the salaries of government workers and spending on public works.

government purchases spending on goods and services by local, state and national governments

The meaning of 'government purchases' requires a little clarification. When the government pays the salary of an army general, that salary is part of government purchases. But what happens when the government pays a social security benefit to one of the elderly? Such government spending is called a **transfer payment** because it is not made in exchange for a currently produced good or service. Transfer payments alter household income, but they do not reflect the economy's production. (From a macroeconomic standpoint, transfer payments are like negative taxes.) Because GDP is intended to measure income from, and expenditure on, the production of goods and services, transfer payments are not counted as part of government purchases.

transfer payment a payment for which no good or service is exchanged

Net Exports

Net exports equal the purchases of domestically produced goods and services by foreigners (exports) minus the domestic purchases of foreign goods (imports). A domestic firm's sale to a buyer in another country, such as the sale of Sony phones to customers in the USA, increases Japanese net exports.

CASE STUDY Revisions to GDP Figures

National or regional statistical services such as the ONS or Eurostat have very sophisticated systems for collecting and calculating national income data. The European system of national and regional accounts (known as ESA95) compiles data on the structure and developments of member state economies. It uses an agreed accounting framework which is recognized across the world to allow each economy to be described and to be able to draw accurate and reliable comparisons between different regions and economies or groups of economies. It is worth trying to imagine how challenging a task collecting all this data actually is. Just look around you whenever you are out at the amount of economic activity which is taking place – people buying and selling goods and services in shops, online transactions, people travelling on public transport, people using their cars for which they have bought insurance, fuel and paid for maintenance, road sweepers cleaning up litter, emptying bins and removing chewing gum off the pavements, police patrolling the streets and so on. Trying to collect all the data for the expenditures that are taking place is a massive task. Inevitably, any published GDP figure is provisional and statistical services continue to revise the data they receive and publish revised figures on a regular basis.

Providing accurate data is important because business, government and investment decisions are all made with the macroeconomic environment in mind. It is made very clear, therefore, that GDP data when published is subject to revision and these revisions can sometimes be significant. In January 2013, for example, the ONS published provisional GDP statistics for the fourth quarter of 2012. The third quarter figures had provided some good news amidst a year which was generally depressing for the UK. Third quarter GDP was reported as increasing by 0.9 per cent largely due to the effect of the Olympic Games which the UK hosted in the summer of 2012. The fourth quarter figures showed the economy had shrunk by 0.3 per cent. However, the ONS reported that the figures, which are often called 'flash estimates', were based on just 44 per cent of the total data used for final estimate. There will be at least two revisions to this data and it may be that once all the data is in, the 0.3 per cent decrease will be confirmed or revised up or down. Quarter 2 data, for example, was originally published as a contraction of 0.7 per cent in April but then revised to -0.5 per cent and then to -0.4 per cent when all the data were available in September 2012. Announcements on GDP data, therefore, needs to be looked at with some caution – one quarter's figures may not necessarily be a reliable indicator of the pattern of GDP over time.



The complexity of capturing data from all the transactions and trade which takes place in an economy is such that revision to GDP data is a fact of life.

net exports spending on domestically produced goods and services by foreigners (exports) minus spending on foreign goods by domestic residents (imports)

The 'net' in 'net exports' refers to the fact that the value of imports is subtracted from the value of exports. This subtraction is made because imports of goods and services are included in other components of GDP. For example, suppose that a UK household buys a £30,000 car from Volvo, the Swedish car maker. That transaction increases consumption in the UK by £30,000 because car purchases are part of consumer spending in the UK. It also reduces net exports by £30,000 because the car is an import (note it represents an export for Sweden). In other words, net exports include goods and services produced abroad (with a minus sign) because these goods and services are included in consumption, investment and government purchases (with a plus sign). Thus, when a domestic household, firm or government buys a good or service from abroad, the purchase reduces net exports – but because it also raises consumption, investment or government purchases, it does not affect GDP. The above example shows the importance of making sure that we focus on a particular country when discussing imports and exports because of the potential for confusion to arise.

Once GDP figures are published they can be presented in different ways. One of the most common is GDP per capita. **GDP per capita** is found by dividing the GDP of a country by the population of that country to express national income per head of the population. This measure is useful in comparing GDP across different countries.

GDP per capita gross domestic product divided by the population of a country to give a measure of national income per head

SELF TEST List the four components of expenditure. Which of these do you think accounts for the largest proportion of GDP in a country? Why?

REAL VERSUS NOMINAL GDP

As we have seen, GDP measures the total expenditure on goods and services in all markets in the economy. If total spending rises from one year to the next, one of two things must be true (or a combination of the two): (1) the economy is producing a larger output of goods and services (a real increase); or (2) goods and services are being sold at higher prices (a nominal increase). When studying changes in the economy over time, economists want to separate these two effects. In particular, they want a measure of the total quantity of goods and services the economy is producing that is not affected by changes in the prices of those goods and services.

To do this, economists use a measure called **real GDP**. Real GDP answers a hypothetical question: what would be the value of the goods and services produced this year if we valued these goods and services at the prices that prevailed in some specific year in the past? By evaluating current production using prices that are fixed at past levels, real GDP shows how the economy's overall production of goods and services changes over time. The GDP figures produced using this method are called **GDP at constant prices**. GDP figures produced without taking into consideration the change in prices over time are called **GDP at current or market prices** and are calculated by taking the output of goods and services and multiplying by the price of those goods and services in the reporting year.

real GDP the measure of the value of output in the economy which takes into account changes in prices over time
GDP at constant prices gross domestic product calculated using prices that existed at a particular base year which takes into account changes in inflation over time
GDP at current or market prices gross domestic product calculated by multiplying the output of goods and services by the price of those goods and services in the reporting year

To see more precisely how real GDP is constructed, let's consider an example.

A Numerical Example

Table 20.1 shows some data for an economy that produces only two goods – apples and potatoes. The table shows the quantities of the two goods produced and their prices in the years 2013, 2014 and 2015.

To compute total spending in this economy, we would multiply the quantities of apples and potatoes by their prices. In the year 2013, 100 kg of apples are sold at a price of €1 per kg, so expenditure on apples equals €100. In the same year, 50 kg of potatoes are sold for €2 per kg, so expenditure on potatoes also equals €100. Total expenditure in the economy – the sum of expenditure on apples and expenditure on potatoes – is €200. This amount, the production of goods and services valued at current or market prices (i.e. the price existing in the reporting year), is called **nominal GDP**.

nominal GDP the production of goods and services valued at current prices

TABLE 20.1 Real and Nominal GDP

This table shows how to calculate real GDP, nominal GDP and the GDP deflator for a hypothetical economy that produces only apples and potatoes.

| Prices and quantities | | | | |
|---------------------------------------|--|-------------------------|------------------------------|---------------------------|
| Year | Price of apples per kg (€) | Quantity of apples (kg) | Price of potatoes per kg (€) | Quantity of potatoes (kg) |
| 2013 | 1 | 100 | 2 | 50 |
| 2014 | 2 | 150 | 3 | 100 |
| 2015 | 3 | 200 | 4 | 150 |
| Calculating nominal GDP | | | | |
| Year | | | | |
| 2013 | $(€1 \text{ per kg apples} \times 100 \text{ kg}) + (€2 \text{ per kg potatoes} \times 50 \text{ kg}) = €200$ | | | |
| 2014 | $(€2 \text{ per kg apples} \times 150 \text{ kg}) + (€3 \text{ per kg potatoes} \times 100 \text{ kg}) = €600$ | | | |
| 2015 | $(€3 \text{ per kg apples} \times 200 \text{ kg}) + (€4 \text{ per kg potatoes} \times 150 \text{ kg}) = €1,200$ | | | |
| Calculating real GDP (base year 2013) | | | | |
| Year | | | | |
| 2013 | $(€1 \text{ per kg apples} \times 100 \text{ kg}) + (€2 \text{ per kg potatoes} \times 50 \text{ kg}) = €200$ | | | |
| 2014 | $(€1 \text{ per kg apples} \times 150 \text{ kg}) + (€2 \text{ per kg potatoes} \times 100 \text{ kg}) = €350$ | | | |
| 2015 | $(€1 \text{ per kg apples} \times 200 \text{ kg}) + (€2 \text{ per kg potatoes} \times 150 \text{ kg}) = €500$ | | | |
| Calculating the GDP deflator | | | | |
| Year | | | | |
| 2013 | $(€200/€200) \times 100 = 100$ | | | |
| 2014 | $(€600/€350) \times 100 = 171$ | | | |
| 2015 | $(€1,200/€500) \times 100 = 240$ | | | |

The table shows the calculation of nominal GDP for these three years. Total spending rises from €200 in 2013 to €600 in 2014 and then to €1,200 in 2015. Part of this rise is attributable to the increase in the quantities of apples and potatoes, and part is attributable to the increase in the prices of apples and potatoes.

To obtain a measure of the amount produced that is not affected by changes in prices, we use real GDP, which is the production of goods and services valued at constant prices. We calculate real GDP by first choosing one year as a *base year*. We then use the prices of apples and potatoes in the base year to compute the value of goods and services in all of the years. In other words, the prices in the base year provide the basis for comparing quantities in different years.

Suppose that we choose 2013 to be the base year in our example. We can then use the prices of apples and potatoes in 2013 to compute the value of goods and services produced in 2013, 2014 and 2015. Table 20.1 shows these calculations. To compute real GDP for 2013 we use the prices of apples and potatoes in 2013 (the base year) and the quantities of apples and potatoes produced in 2013. (Thus, for the base year, real GDP always equals nominal GDP.) To compute real GDP for 2014, we use the prices of apples and potatoes in 2013 (the base year) and the quantities of apples and potatoes produced in 2014. Similarly, to compute real GDP for 2015, we use the prices in 2013 and the quantities in 2015. When we find that real GDP has risen from €200 in 2013 to €350 in 2014 and then to €500 in 2015, we know that the increase is attributable to an increase in the *quantities produced*, because the prices are being held fixed at base-year levels.

To sum up: nominal GDP uses current or market prices to place a value on the economy's production of goods and services, while real GDP uses constant base-year prices to place a value on the economy's production of goods and services. Because real GDP is not affected by changes in prices, changes in real GDP reflect only changes in the amounts being produced. Thus, real GDP is a measure of the economy's production of goods and services.

Our goal in computing GDP is to gauge how well the overall economy is performing. Because real GDP measures the economy's production of goods and services, it reflects the economy's ability to satisfy people's needs and desires. Thus, real GDP is a better gauge of economic well-being than is nominal GDP. When economists talk about the economy's GDP, they usually mean real GDP rather than nominal GDP. And when they talk about growth in the economy, they measure that growth as the percentage change in real GDP from one period to another.

The GDP Deflator

As we have just seen, nominal GDP reflects both the prices of goods and services and the quantities of goods and services the economy is producing. In contrast, by holding prices constant at base-year levels, real GDP reflects only the quantities produced. From these two statistics we can compute a third, called the **GDP deflator**, which reflects the prices of goods and services but not the quantities produced.

GDP deflator a measure of the price level calculated as the ratio of nominal GDP to real GDP times 100

The GDP deflator is calculated as follows:

$$\text{GDP deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}} \times 100$$

Because nominal GDP and real GDP must be the same in the base year, the GDP deflator for the base year always equals 100. The GDP deflator for subsequent years measures the change in nominal GDP from the base year that cannot be attributable to a change in real GDP.

The GDP deflator measures the current level of prices relative to the level of prices in the base year. To see why this is true, consider a couple of simple examples. First, imagine that the quantities produced in the economy rise over time but prices remain the same. In this case, both nominal and real GDP rise together, so the GDP deflator is constant. Now suppose, instead, that prices rise over time but the quantities produced stay the same. In this second case, nominal GDP rises but real GDP remains the same, so the GDP deflator rises as well. Notice that, in both cases, the GDP deflator reflects what's happening to prices, not quantities.

Let's now return to our numerical example in Table 20.1. The GDP deflator is computed at the bottom of the table. For year 2013, nominal GDP is €200, and real GDP is €200, so the GDP deflator is 100. For the year 2014, nominal GDP is €600, and real GDP is €350, so the GDP deflator is 171. Because the GDP deflator rose in year 2014 from 100 to 171, we can say that the price level increased by 71 per cent.

The GDP deflator is one measure that economists use to monitor the average level of prices in the economy. We examine another – the consumer prices index – in the next chapter, where we also describe the differences between the two measures.

SELF TEST Define real and nominal GDP. Which is a better measure of economic well-being? Why?

FYI



Measuring GDP

We have looked at measuring GDP through income and expenditure methods and said that in theory they are the same. GDP can also be measured by looking at the output method. The price of a product represents the value of the inputs (land, labour and capital) that went into production. At each stage of the production process the value or output can be recorded – the value added. GDP is also a measure of the value of output in the economy. Statistics offices usually classify output according to industry types – construction, agriculture forestry and fisheries, mining, services and so on.

The measure for GDP is essentially what is called a 'volume measure'. In other words, it is measuring 'how much' has been produced. In the UK the ONS has to produce a single measure of GDP and does this by using the three approaches – income, expenditure and output – and balancing the final figure in relation to the three measures. Remember, we said that in theory they should all be the same but the process of collecting information from millions of transactions means that in reality they would never be exact.

Imagine a situation, for example, where someone has some work done on his or her house – say some

plastering of a room. The total price for the work paid by the owner of the house might be £1,500 – this should represent the sum of the resources used in the job. The plasterer may declare to Her Majesty's Revenue and Customs (HMRC), that is responsible for collecting UK taxes, that the work represented income of £1,200 (to evade paying income tax – illegal, but nevertheless it happens!) and the resources used might have included the plasterboard, plaster, electricity, buckets, nails, water, trowels and hawks, the vehicles to get to and from the builders' merchants, etc. It is easy to see how, even in such a simple example, it becomes incredibly difficult to keep track of every item involved!

The choice of which base year to use in calculating real GDP is also important. Up until early 2004, the ONS used a method referred to as 'fixed base aggregations'. For example, they might have compared the data using 1995 as the base year. However, as with any statistical data which use a base year in their compilation, this can lead to inconsistencies in the data because circumstances change. Goods, for example, that were common in 1995 may not exist anymore and there might be some products that were not available at that time – DVDs,

for example. In general, the more you can update the base year, the more accurate the statistics will be. Fixed base aggregations in GDP statistics were used as follows. The growth of the different parts of the economy was given weights when trying to arrive at the final figure for GDP. This takes account of their relative importance in the economy and a base year of 1995 was used to help determine these weights. These base years were updated every five years but, as we have seen, this can lead to problems.

The method known as 'annual chain linking' helps to overcome these problems. Rather than updating the base year every five years, this method does it every year, calculating the prices in previous years' prices (PYPs). You will see GDP tables expressed as 'GDP (CVM)', where CVM means Chained Volume Measures, as opposed to 'GDP in constant prices'. The ONS and other statistical services are constantly looking to update their methods to provide more reliable and accurate data and this is one example of such a revision of methods.

In terms of analysis of the data it should make little difference, but recognizing it enables you to be more confident that the data are more likely to be accurate and reliable compared to the previous measures.

GDP AND ECONOMIC WELL-BEING

Earlier in this chapter, GDP was referred to as the best single measure of the economic well-being of a society. Now that we know what GDP is, we can evaluate this claim.

As we have seen, GDP measures both the economy's total income and the economy's total expenditure on goods and services. Thus, GDP per person tells us the income and expenditure of the average person in the economy. Because most people would prefer to receive higher income and enjoy higher expenditure, GDP per person seems a natural measure of the economic well-being of the average individual.

Yet some people dispute the validity of GDP as a measure of well-being. Surely we should not be obsessed with material things and income? Surely we should value things that are not measured by GDP and that contribute to the quality of life and economic well-being, such as the health of a country's children, or the quality of their education, or even the beauty of the poetry making up our literary heritage?

The answer is that a large GDP does in fact help us to lead a good life. GDP does not measure the health of our children, but nations with larger GDP can afford better health care for their children. GDP does not measure the quality of their education, but nations with larger GDP can afford better educational systems. GDP does not measure the beauty of our poetry, but nations with larger GDP can afford to teach more of their citizens to read and enjoy poetry. GDP does not take account of our intelligence, integrity, courage or wisdom, but all of these laudable attributes are easier to foster when people are less concerned about being able to afford the material necessities of life. In short, GDP does not directly measure those things that make life worthwhile, but it does measure our ability to obtain the inputs into a worthwhile life.

GDP is not, however, a perfect measure of well-being. Some things that contribute to a good life are left out of GDP. One is leisure. Suppose, for instance, that everyone in the economy suddenly started working every day of the week, rather than enjoying leisure on weekends. More goods and services would be produced, and GDP would rise. Yet, despite the increase in GDP, we should not conclude that everyone would be better off. The loss from reduced leisure would offset the gain from producing and consuming a greater quantity of goods and services.

Because GDP uses market prices to value goods and services, it excludes the value of almost all activity that takes place outside of markets. In particular, GDP omits the value of goods and services produced at home. When a chef prepares a delicious meal and sells it at his restaurant, the value of that meal is part of GDP. But if the chef prepares the same meal for his wife, the value he has added to the raw ingredients is left out of GDP. As we saw above, child care provided in day care centres is part of GDP, whereas child care by parents at home is not. Volunteer work also contributes to the well-being of those in society, but GDP does not reflect these contributions.

Another thing that GDP excludes is the quality of the environment. Imagine that the government eliminated all environmental regulations. Firms could then produce goods and services without considering the pollution they create, and GDP might rise. Yet well-being would most likely fall. The deterioration in the quality of air and water would more than offset the gains from greater production.

GDP also says nothing about the distribution of income. A society in which 100 people have annual incomes of €50,000 has GDP of €5 million and, not surprisingly, GDP per capita (person) of €50,000; so does a society in which 10 people earn €500,000 and 90 suffer with nothing at all. Few people would look at those two situations and call them equivalent. GDP per person tells us what happens to the average person, but behind the average lies a large variety of personal experiences.

In the end, we can conclude that GDP is a good measure of economic well-being for most – but not all – purposes. It is important to keep in mind what GDP includes and what it leaves out.

The Economics of Happiness

It does appear that, despite the massive increase in wealth, incomes and access to material goods and services for many people in developed countries over the last 50 years, our perception of happiness has not really changed that much. Increased wealth has not brought with it similar increases in happiness. Numerous surveys have highlighted relatively stable rates of 'happiness' in rich countries. Professor Richard Layard, one of a group of economists including Andrew Oswald, Stephen Nickell, Robert Skidelsky, Tim Besley, Will Hutton and those behavioural economists that we have met before in this book, have

studied this apparent paradox. Layard states that in relation to Western societies: 'on average, people are no happier than they were fifty years ago' (see Layard, R. (2005) *Happiness: Lessons from a New Science*. London: Penguin).

Many people perceive January to be a very stressful month. In 2006, a number of newspapers and media organizations ran a story suggesting that 23 January 2006 was going to be the official 'worst day of the year'. How was this arrived at? It appears that someone called Cliff Arnall had devised a formula for calculating it. The formula is: $[W + (D - d)] \times TQM \times NA$, where W is the weather, D is debt, T is time since Christmas, Q the time since failing to keep to our New Year's resolutions, M is motivational levels and NA the need to take action.

This formula was an attempt to quantify a way in which states of happiness (or sadness) can be measured. Its scientific basis has been called into question. Arnall said that it is useful in triggering discussion and debate about what we understand by happiness and to think of ways we can deal with the effects of how we live our lives on our well-being.

What is needed, therefore, is a widely accepted set of criteria that is accepted as constituting happiness. The increased understanding of the brain and neuroscience has helped to shed some light on this but it is a complex area. Psychologists and economists have found that links between how people see their own happiness and other things that might influence, such as judgement, are statistically strong. This also seems to apply across different countries. As a result, we might be able to conclude that there are a range of factors that can contribute to a definition of happiness; that if you are fortunate enough to find yourself being able to boast having these characteristics, or in some cases managing to avoid them, then you are more likely to be happy.

Layard identified some key factors that may contribute to 'happiness'. He suggested that some of the factors associated with promoting happiness, include sex, socializing, relaxing, praying, worshipping or meditating, eating, exercising, watching TV and shopping, among others. Other studies have suggested that an individual's level of education, health, whether they are married, single or divorced, the level of income enjoyed, whether they are working/unemployed/retired, their aspirations and whether they have experienced bereavement can all be contributory factors.

From such characteristics, equations to statistically arrive at measures of well-being can be derived. These have been used by both economists and psychologists and have been found to be surprisingly reliable in statistical terms. One of the leading thinkers on the economics of happiness is Professor Andrew Oswald who is based at the University of Warwick. Oswald offers the following formula:

$$W_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$$

In this formula, W_{it} is the reported well-being of an individual at a particular time period, X represents a collection of variables that are known to be characteristics affecting well-being at a particular time period: these could be economic, such as income, or demographic, such as gender, with various points in between. The final term is an error term, which is used to take into account unobserved factors that may have an impact on the final outcome.

One of the factors above that may be affecting happiness levels more than we might expect is aspirations. You are very likely, whatever your age, to have been regaled by your parents that things were different in their day and that people these days have so much more. That is very true but what might also be the case is that different generations are starting to expect more. If your parents have managed to reach middle age and have a comfortable BMW to drive around in, you might start to expect such a vehicle to be the norm and would hope it might be your first car and so on.

Layard suggests that unhappiness results from society viewing how it is developing in terms of a zero-sum game. What he means by this is that we might increasingly view the scramble to gain money and status in terms of a competitive game that has a winner and a loser. If I get a high ranking job with a large salary and lots of status, it means that you somehow lose out – either by not being able to get that same job or in some sort of psychological way. Such a perception of life being a zero-sum game is a source of much unhappiness.

If we know something of the factors that can contribute to making someone 'happy' and also that our current measure is not the best at reflecting this, then it makes sense to look elsewhere for a measure of well-being. One such idea is to use something called the Measure of Domestic Progress (MDP). The MDP

looks at many of the factors that we might associate with economic growth but takes into consideration the relative effects of economic growth and factors in other things that GDP calculations do not consider. For example, there is an attempt to place a value on the amount of unpaid domestic work that is carried out, which is taken as being a positive factor in contributing to well-being. It also assigns a negative effect to various social and environmental impacts of growth such as pollution, depletion of natural resources and costs of crime and family breakdown.

Other attempts to suggest more effective measures of improvements in well-being include such things as the Measure of Economic Welfare (MEW) developed by James Tobin and William Nordhaus, the Index of Sustainable Economic Welfare and the Genuine Progress Indicator.

It is proving difficult for any of these alternative measures to make a breakthrough and to be accepted in widespread use as a measure of well-being. The problems of attempting to measure subjective human behaviour and expose it to rigorous science are once again proving a challenge to economists. We might think that having some form of 'happiness index' would be a good thing, but if it does not really tell us what we want it to tell us, it is not very useful. There again, GDP figures do not seem to be telling us the whole story either!

International Differences in GDP and the Quality of Life

One way to gauge the usefulness of GDP as a measure of economic well-being is to examine international data. Rich and poor countries have vastly different levels of GDP per person. If a large GDP leads to a higher standard of living, then we should observe GDP to be strongly correlated with measures of the quality of life. And, in fact, we do.

Table 20.2 shows 13 of the world's countries ranked in order of GDP per person. The table also shows life expectancy (the expected life span at birth) and literacy (the percentage of the adult population who can read). These data show a clear pattern. In rich countries, such as the United Kingdom, the United States, Japan and Germany, people can expect to live into their late seventies or early eighties, and almost all of the population can read. In poor countries, such as Niger, Bangladesh and Pakistan, people typically can expect to live much shorter lives and the proportion of the population which is literate is relatively low.

TABLE 20.2 GDP, Life Expectancy and Literacy

The table shows GDP per person (measured in US dollars) and two measures of the quality of life for 13 countries.

| Country | GDP per capita (Current US\$) (2011) | Life expectancy at birth (males and females) total years (2011) | Literacy rate (% of people over 15), adult total (2011) |
|----------------|--|--|---|
| United States | 48,112 | 78 (2010) | 99 |
| Japan | 45,903 | 83 | 99 |
| Germany | 44,060 | 80 (2010) | 99 |
| United Kingdom | 39,038 | 80 | 99 |
| Russia | 13,089 | 69 (2010) | 100 |
| Mexico | 10,047 | 77 (2010) | 93 (2010) |
| Brazil | 12,594 | 73 | 90 (2009) |
| China | 5,445 | 73 (2010) | 94 (2010) |
| Indonesia | 3,495 | 69 | 92 (2010) |
| India | 1,489 | 65 | 63 (2010) |
| Pakistan | 1,189 | 65 (2010) | 56 (2010) |
| Bangladesh | 743 | 69 | 56 (2010) |
| Niger | 374 | 55 | 29 |

Although data on other aspects of the quality of life are less complete, they tell a similar story. Countries with low GDP per person tend to have more infants with low birth weight, higher rates of infant mortality, higher rates of maternal mortality, higher rates of child malnutrition and less common access to safe drinking water. In countries with low GDP per person, fewer school-age children are actually in school, and those who are in school must learn with fewer teachers per student.

These countries also tend to have fewer televisions, fewer telephones, fewer paved roads and fewer households with electricity. International data leave no doubt that a nation's GDP is closely associated with its citizens' standard of living.

SELF TEST Why should policymakers care about GDP?

CONCLUSION

This chapter has discussed how economists measure the total income of a nation. Measurement is, of course, only a starting point. Much of macroeconomics is aimed at revealing the long-run and short-run determinants of a nation's gross domestic product. Why, for example, is GDP higher in the United Kingdom and Japan than in India and Nigeria? What can the governments of the poorest countries do to promote more rapid growth in GDP? Why does GDP in European and North American countries rise rapidly in some years and fall in others? What can policymakers do to reduce the severity of these fluctuations in GDP? These are the questions we will take up shortly.

At this point, it is important to acknowledge the importance of just measuring GDP. We all get some sense of how the economy is doing as we go about our lives. But the economists who study changes in the economy and the policymakers who formulate economic policies need more than this vague sense – they need concrete data on which to base their judgements. Quantifying the behaviour of the economy with statistics such as GDP is, therefore, the first step to developing a science of macroeconomics.

IN THE NEWS



Recession and Recovery

A number of countries in Europe have faced significant economic difficulties in the wake of the financial crisis in 2007–2008. Economic growth in many countries since the crisis has been sluggish with many countries failing to get back to a position they were in prior to the crisis.

Comparing Recessions

At the time of the financial crisis a number of economists predicted that recovery would be long and difficult. Those predictions have proved correct but maybe many people have been surprised by the length of time which has elapsed and for some countries things do not look any brighter. In Greece, for example, the recession which began in 2008 shows no sign of abating and in 2012 the government had to announce that GDP would shrink by

7 per cent in 2012, 2 per cent more than was originally forecast and that there would be no growth until 2014 at the earliest.

In 2009, the 17 Eurozone countries and the UK experienced negative growth which occurs when the value of output is lower than that in the previous time period. In Estonia the economy shrank by over 14 per cent, in Slovenia by 8 per cent, by 8.4 per cent in Finland, by 7.0 per cent in Ireland and by 5.5 per cent in Italy and 5.2 per cent in Germany. The other Eurozone

countries also saw a contraction in the economy but less severe than those quoted, ranging between 0 per cent and 4.4 per cent. In 2010, 15 of the 17 countries returned to growth although in many cases it was modest. By 2011 the rate of growth for some had slowed considerably with growth rates of less than 1 per cent occurring in five countries, and three countries experiencing negative growth.

In the UK the time it is taking to get the economy back onto a more long-term growth trend of



Observe the similarities between this 1930s Depression unemployment line and the present day queue outside the Job Centre – recessions throughout history have many similarities.



around 2.0 per cent is taking much longer than many people thought. The problems in the Eurozone have meant that export markets for the UK have shrunk and the Coalition government's austerity measures to try and cut the budget deficit has hit consumer spending. Firms are reluctant to invest in such a depressed climate although many large firms do have lots of spare cash which is adding to the failure of the UK economy to break out of recession.

In comparison to other recessions, the UK is taking much longer to recover. There have been four major recessionary periods in recent UK economic history, 1930–1934, 1973–1976, 1979–1983 and 1990–1993. Almost five years on from the start of the recession in 2008, the UK economy is about

3.35 per cent smaller. At the same time, five years following the start of the 1930–1934 recession, the economy was 3.83 per cent bigger; after the start of the 1973–1976 recession the economy was 5.45 per cent bigger; after the 1979–1983 recession, 3.04 per cent bigger; and the economy was 8.54 per cent bigger five years after the start of the 1990–1993 recession (data source, NIESR).

With Spain, Portugal, Ireland, Italy and Greece all experiencing austerity policies, the chances of the Eurozone growing strongly is unlikely. Most economists see minimal growth at best for these economies and whilst parts of the Eurozone remain depressed, the hope for trading out of recession for other economies becomes less.

Questions

- 1** Use an example to show how an economy can 'shrink'.
- 2** What is the difference between a slowdown in GDP and a fall in GDP? Use a suitable example to illustrate your answer.
- 3** What do you think is meant by an 'export-led recovery'?
- 4** GDP is the sum of $C + I + G + X - M$. Using the article, explain how the behaviour of each of these components in the UK is affecting GDP.
- 5** Why do you think the UK economy has failed to recover to at least the size it was prior to the start of the recession in 2008 when in all of the instances of previous recessions noted in the article, the economy has recovered so much quicker after the same length of time?

SUMMARY

- Because every transaction has a buyer and a seller, the total expenditure in the economy must equal the total income in the economy.
- Gross domestic product (GDP) measures an economy's total expenditure on newly produced goods and services and the total income earned from the production of these goods and services. More precisely, GDP is the market value of all final goods and services produced within a country in a given period of time.
- GDP is divided among four components of expenditure: consumption, investment, government purchases and net exports. Consumption includes spending on goods and services by households, with the exception of purchases of new housing. Investment includes spending on new equipment and structures, including households' purchases of new housing. Government purchases include spending on goods and services by local, state and central governments. Net exports equal the value of goods and services produced domestically and sold abroad (exports) minus the value of goods and services produced abroad and sold domestically (imports).
- Nominal GDP uses current prices to value the economy's production of goods and services. Real GDP uses constant base-year prices to value the economy's production of goods and services. The GDP deflator – calculated from the ratio of nominal to real GDP – measures the level of prices in the economy.
- GDP is a good measure of economic well-being because people prefer higher to lower incomes. But it is not a perfect measure of well-being. For example, GDP excludes the value of leisure and the value of a clean environment.

QUESTIONS FOR REVIEW

- 1 Explain why an economy's income must equal its expenditure.
- 2 What is meant by the term GDP per capita and how is it measured?
- 3 Which contributes more to GDP – the production of an economy car or the production of a luxury car? Why?
- 4 A farmer sells wheat to a baker for €2. The baker uses the wheat to make bread, which is sold for €3. What is the total contribution of these transactions to GDP?
- 5 Many years ago, Jamanda paid €500 to put together a record collection. Today she sold her albums at a car boot sale for €100. How does this sale affect current GDP?
- 6 List the four components of GDP. Give an example of each.
- 7 Why do economists use real GDP rather than nominal GDP to gauge economic well-being?
- 8 In the year 2014, the economy produces 100 loaves of bread that sell for €2 each. In the year 2015, the economy produces 200 loaves of bread that sell for €3 each. Calculate nominal GDP, real GDP and the GDP deflator for each year (use 2014 as the base year). By what percentage does each of these three statistics rise from one year to the next?
- 9 Why is it desirable for a country to have a large GDP? Give an example of something that would raise GDP and yet be undesirable.
- 10 What are the limitations of using GDP as a measure of well-being for a country?

PROBLEMS AND APPLICATIONS

- 1 What components of GDP (if any) would each of the following transactions affect? Explain.
 - a. A family buys a new refrigerator.
 - b. Aunt Jane buys a new house.
 - c. Aston Martin sells a DB7 from its inventory.
 - d. You buy a pizza.
 - e. The government builds a new motorway.
 - f. You buy a bottle of Californian wine.
 - g. Honda expands its factory in Derby, England.
- 2 The 'government purchases' component of GDP does not include spending on transfer payments such as social security. Thinking about the definition of GDP, explain why transfer payments are excluded.
- 3 Why do you think households' purchases of new housing are included in the investment component of GDP rather than the consumption component? Can you think of a reason why households' purchases of

new cars should also be included in investment rather than in consumption? To what other consumption goods might this logic apply?

- 4 As the chapter states, GDP does not include the value of used goods that are resold. Why would including such transactions make GDP a less informative measure of economic well-being?
- 5 Below are some data from the land of milk and honey.

| Year | Price of milk (€) | Quantity of milk (quarts) | Price of honey (€) | Quantity of honey (quarts) |
|------|-------------------|---------------------------|--------------------|----------------------------|
| 2013 | 1 | 100 | 2 | 50 |
| 2014 | 1 | 200 | 2 | 100 |
| 2015 | 2 | 200 | 4 | 100 |

- a. Compute nominal GDP, real GDP and the GDP deflator for each year, using 2013 as the base year.
 - b. Compute the percentage change in nominal GDP, real GDP and the GDP deflator in 2014 and 2015 from the preceding year. For each year, identify the variable that does not change. Explain in words why your answer makes sense.
 - c. Did economic well-being rise more in 2014 or 2015? Explain.
- 6 If prices rise, people's income from selling goods increases. The growth of real GDP ignores this gain, however. Why, then, do economists prefer real GDP as a measure of economic well-being?
 - 7 One day Boris the Barber, plc, collects €400 for haircuts. Over this day, his equipment depreciates in value by €50. Of the remaining €350, Boris sends €30 to the government in sales taxes, takes home €220 in wages, and retains €100 in his business to add new equipment in the future. From the €220 that Boris takes home, he pays €70 in income taxes. Based on this information, compute Boris' contribution to the following measures of income:
 - a. gross domestic product
 - b. net national product
 - c. national income
 - d. personal income
 - e. disposable personal income.
 - 8 Goods and services that are not sold in markets, such as food produced and consumed at home, are generally not included in GDP. Can you think of how this might cause the numbers in the second column of Table 20.2 to be misleading in a comparison of the economic well-being of the United Kingdom and India? Explain.
 - 9 Economists sometimes prefer to use GNP rather than GDP as a measure of economic well-being. Which measure should we prefer if we are analysing the total income of domestic residents? Which measure should we prefer if we are analysing the total amount of economic activity occurring in the economy?
 - 10 The participation of women in many European and North American economies has risen dramatically over the past three decades.
 - a. How do you think this rise affected GDP?
 - b. Now imagine a measure of well-being that includes time spent working in the home and taking leisure. How would the change in this measure of well-being compare to the change in GDP?
 - c. Can you think of other aspects of well-being that are associated with the rise in women's labour-force participation? Would it be practical to construct a measure of well-being that includes these aspects?

21

MEASURING THE COST OF LIVING

The UK parliament is often referred to as the ‘mother of parliaments’, by which is meant that it is one of the oldest parliaments and one that has served as a model, to a greater or lesser extent, when parliaments have been established in other countries. The origins of the British parliament, as a meeting place where representatives of the various regions and boroughs throughout the country could influence national policy, can be traced as far back as the early 13th century. It is therefore perhaps strange that members of parliament were paid nothing at all for their trouble until 1911, when they first received a salary of £400 (€460) a year. This amount was decreased to £360 (€414) in 1931, but has generally increased over time, so that, for example, MPs received £600 (€690) in 1937, £6,270 (€7,214) in 1977 and £65,738 (€75,644) in 2012. The 2012 pay doesn’t sound too bad (especially as this excludes various allowances to which the MP is also entitled at this time) – it is, after all, over two and a half times what the average bus driver in the UK earns – around £24,000 (€27,616) and double what the average secondary school teacher earns – around £32,000 (€36,822). But the pay of £400 a year – about £8 a week (€9.20) in 1911 seems pretty poor, and it even went down by 10 per cent over the ensuing 20 years.

But, as everyone knows, the prices of goods and services have also changed over time and have for the most part increased. In 1900, a copy of *The Times* was priced at 1.2p, a pint of beer and milk could be bought for less than 1p and a dozen fresh eggs was 6.9p. In 1930, the average house price was £590; a loaf of bread was around 3p (These prices are expressed in decimal currency equivalents). At the turn of the 21st century, the average level of prices was 66 times the level of 1900. Because prices were so much lower in the early 1900s than they are today, it is not clear whether an MP in 1911 enjoyed a higher or lower standard of living than today’s MPs.

In the preceding chapter we looked at how economists use gross domestic product (GDP) to measure the quantity of goods and services that the economy is producing. This chapter examines how economists measure the overall cost of living. To compare an MP’s salary of £400 in 1911 to salaries from today, we need to find some way of turning money figures into meaningful measures of purchasing power. That is exactly the job of a statistic called the *consumer prices index*. After seeing how the consumer prices index is constructed, we discuss how we can use such a prices index to compare money figures from different points in time.

The consumer prices index is used to monitor changes in the cost of living over time. When the consumer prices index rises, the typical family has to spend more money to maintain the same standard of living. Economists use the term *inflation* to describe a situation in which the economy’s overall price level is rising. The *inflation rate* is the percentage change in the price level from the previous period. As we will see in the coming chapters, inflation is a closely watched aspect of macroeconomic performance and is a key variable guiding macroeconomic policy. This chapter provides the background for that analysis by showing how economists measure the inflation rate using the consumer prices index.

THE CONSUMER PRICES INDEX

The **consumer prices index** (CPI) is a measure of the overall prices of the goods and services bought by a typical consumer. Each month, a government bureau – in the UK the Office for National Statistics (ONS) and in Europe, Eurostat – computes and reports the CPI. In this section we discuss how the CPI

is calculated and what problems arise in its measurement. We also consider how this index compares to the GDP deflator, another measure of the overall level of prices, which we examined in the preceding chapter.

consumer prices index a measure of the overall prices of the goods and services bought by a typical consumer

How the Consumer Prices Index Is Calculated

When the ONS or Eurostat calculates the CPI and the inflation rate, it uses data on the prices of thousands of goods and services. To see exactly how these statistics are constructed, let's consider a simple economy in which consumers buy only two goods – hot dogs and burgers. Table 21.1 shows the five steps that the ONS and Eurostat follow. (We will use the ONS as the base for the example here but the principle applies to the way price changes are measured in Europe as a whole.)

TABLE 21.1 Calculating the Consumer Prices Index and the Inflation Rate: An Example

This table shows how to calculate the CPI and the inflation rate for a hypothetical economy in which consumers buy only hot dogs and burgers.

| Step 1: Survey consumers to determine a fixed basket of goods | | |
|---|---|-----------------------------|
| 4 hot dogs, 2 burgers | | |
| Step 2: Find the price of each good in each year | | |
| Year | Price of hot dogs (€) | Price of burgers (€) |
| 2014 | 1 | 2 |
| 2015 | 2 | 3 |
| 2016 | 3 | 4 |
| Step 3: Compute the cost of the basket of goods in each year | | |
| 2014 | $(€1 \text{ per hot dog} \times 4 \text{ hot dogs}) + (€2 \text{ per burger} \times 2 \text{ burgers}) = €8$ | |
| 2015 | $(€2 \text{ per hot dog} \times 4 \text{ hot dogs}) + (€3 \text{ per burger} \times 2 \text{ burgers}) = €14$ | |
| 2016 | $(€3 \text{ per hot dog} \times 4 \text{ hot dogs}) + (€4 \text{ per burger} \times 2 \text{ burgers}) = €20$ | |
| Step 4: Choose one year as a base year (2014) and compute the consumer prices index in each year | | |
| 2014 | $(€8/€8) \times 100 = 100$ | |
| 2015 | $(€14/€8) \times 100 = 175$ | |
| 2016 | $(€20/€8) \times 100 = 250$ | |
| Step 5: Use the consumer prices index to compute the inflation rate from previous year | | |
| 2015 | $(175 - 100)/100 \times 100 = 75\%$ | |
| 2016 | $(250 - 175)/175 \times 100 = 43\%$ | |

- 1. Fix the basket.** The first step in computing the consumer prices index is to determine which prices are most important to the typical consumer. If the typical consumer buys more hot dogs than burgers, then the price of hot dogs is more important than the price of burgers and, therefore, should be given greater weight in measuring the cost of living. The ONS sets these weights by surveying consumers and finding the basket of goods and services that the typical consumer buys. In the example in Table 21.1, the typical consumer buys a basket of 4 hot dogs and 2 burgers.

2. *Find the prices.* The second step in computing the CPI is to find the prices of each of the goods and services in the basket for each point in time. The table shows the prices of hot dogs and burgers for three different years.
3. *Compute the basket's cost.* The third step is to use the data on prices to calculate the cost of the basket of goods and services at different times. The table shows this calculation for each of the three years. Notice that only the prices in this calculation change. By keeping the basket of goods the same (4 hot dogs and 2 burgers), we are isolating the effects of price changes from the effect of any quantity changes that might be occurring at the same time.
4. *Choose a base year and compute the index.* The fourth step is to designate one year as the base year, which is the benchmark against which other years are compared. To calculate the index, the price of the basket of goods and services in each year is divided by the price of the basket in the base year, and this ratio is then multiplied by 100. The resulting number is the CPI.

In the example in the table, the year 2014 is the base year. In this year, the basket of hot dogs and burgers costs €8. Therefore, the price of the basket in all years is divided by €8 and multiplied by 100. The CPI is 100 in 2014. (The index is always 100 in the base year.) The CPI is 175 in 2015. This means that the price of the basket in 2015 is 175 per cent of its price in the base year. Put differently, a basket of goods that costs €100 in the base year costs €175 in 2015. Similarly, the consumer prices index is 250 in 2016, indicating that the price level in 2016 is 250 per cent of the price level in the base year.

5. *Compute the inflation rate.* The fifth and final step is to use the CPI to calculate the **inflation rate**, which is the percentage change in the price index from the preceding period. That is, the inflation rate between two consecutive years is computed as follows:

$$\text{Inflation rate in year 2} = 100 \times \frac{(\text{CPI in year 2} - \text{CPI in year 1})}{\text{CPI in year 1}}$$

inflation rate the percentage change in the price index from the preceding period

In our example, the inflation rate is 75 per cent in 2015 and 43 per cent in 2016.

Although this example simplifies the real world by including only two goods, it shows how the ONS computes the CPI and the inflation rate. The ONS collects and processes data on the prices of thousands of goods and services every month and, by following the five foregoing steps, determines how quickly the cost of living for the typical consumer is rising. When the ONS makes its monthly announcement of the CPI, you can usually hear the number on the evening television news or see it in the next day's newspaper.

In addition to the CPI for the overall economy, the ONS also calculates price indices for the sub-categories of 'goods' and of 'services' separately, as well as the **producer prices index** (PPI), which measures the prices of a basket of goods and services bought by firms rather than consumers. Because firms eventually pass on their costs to consumers in the form of higher consumer prices, changes in the PPI are often thought to be useful in predicting changes in the CPI.

producer prices index a measure of the prices of a basket of goods and services bought by firms

Problems in Measuring the Cost of Living

The goal of the CPI is to measure changes in the cost of living – how much people have to pay to purchase goods and services. In other words, the CPI tries to gauge how much incomes must rise in order to maintain a constant standard of living. The standard of living is measured by the amount of goods and services

people are able to buy. Assume that an individual in 2014 has an income of €100 per week and only buys burgers at €2 each. Their standard of living is 50 burgers a week. In 2015, burgers rise in price to €3 and so the standard of living, assuming income does not change, is now only 33.3 burgers a week. We would say that the individual's standard of living has fallen because he can now afford to consume fewer burgers. To keep his standard of living constant at 50 burgers, the person's income needs to rise from €100 per week to €150 per week. If the rise in incomes keeps pace with the rise in prices then the individual's standard of living will remain constant. If incomes rise by a lower percentage than the CPI then standards of living are being eroded and, if incomes rise at a faster rate than the CPI, standards of living are improving and people are better off.

The CPI, however, is not a perfect measure of the cost of living. Three problems with the index are widely acknowledged but difficult to solve.

Substitution Bias The first problem is called *substitution bias*. When prices change from one year to the next, they do not all change proportionately: some prices rise more than others and some prices fall. Consumers respond to these differing price changes by buying less of the goods whose prices have risen by large amounts and by buying more of the goods whose prices have risen less or perhaps even have fallen (and, of course, different consumers will respond to price changes in different ways because of the price elasticity of demand). That is, consumers substitute towards goods that have become relatively less expensive. If a price index is computed assuming a fixed basket of goods, it ignores the possibility of consumer substitution and, therefore, overstates the increase in the cost of living from one year to the next.

Let's consider a simple example. Imagine that in the base year apples are cheaper than pears, and so consumers buy more apples than pears. When the ONS constructs the basket of goods, it will include more apples than pears. Suppose that next year pears are cheaper than apples. Consumers will naturally respond to the price changes by buying more pears and fewer apples. Yet, when computing the CPI, the ONS uses a fixed basket, which in essence assumes that consumers continue buying the now expensive apples in the same quantities as before. For this reason, the index will measure a much larger increase in the cost of living than consumers actually experience.

The Introduction of New Goods The second problem with the CPI is the *introduction of new goods* as highlighted in the following FYI box. When a new good is introduced, consumers have more variety from which to choose. Greater variety, in turn, makes each pound more valuable, so consumers need fewer pounds to maintain any given standard of living. Yet because the CPI is based on a fixed basket of goods and services, it does not reflect this change in the purchasing power of the pound.

Again, let's consider an example. When video players were introduced, consumers were able to watch their favourite films at home. Compared with going to the cinema, the convenience was greater and the cost was less. A perfect cost of living index would have reflected the introduction of the video player with a decrease in the cost of living. The CPI, however, did not decrease in response to the introduction of the video player. Eventually, the ONS did revise the basket of goods to include video players, and subsequently the index reflected changes in their prices. But the reduction in the cost of living associated with the initial introduction of the video player never showed up in the index. As time has passed the index has to be revised to take account of the decline in video and the rise of DVD, then Blu-ray DVD and more recently the increase in downloading of movies.

Unmeasured Quality Change The third problem with the consumer prices index is *unmeasured quality change*. If the quality of a good deteriorates from one year to the next, the effective value of a pound falls, even if the price of the good stays the same. Similarly, if the quality rises from one year to the next, the effective value of a pound rises. The ONS does its best to account for quality change. When the quality of a good in the basket changes – for example, when a car model has more horsepower or gets better petrol mileage from one year to the next – the ONS adjusts the price of the good to account for the quality change. It is, in essence, trying to compute the price of a basket of goods of constant quality.

FYI



What is in the CPI's Basket?

When constructing the CPI, the UK ONS tries to include all the goods and services that the typical consumer buys. Moreover, it tries to weight these goods and services according to how much consumers buy of each item. Every month the ONS collects around 180,000 prices of 700 goods and services from around 150 areas throughout the UK that are supposed to be representative of the goods and services we use on a regular basis. As buying habits change, the basket of goods and services has to change also. Each year the ONS announces a revision to the basket of goods and services that make up the basis of inflation figures in the UK. The ONS has to take into consideration not only how representative the sample of goods and services is but also any changes in importance in the typical household budget and whether the replacement brands that shops bring in are of comparable quality.

Table 21.2 shows the main groupings of the products included in the index. In this case, the index is the Retail Prices Index (RPI) which is constructed in exactly the same way as the CPI but includes some product areas, notably housing costs, which the CPI does not. The UK adopted the CPI as its official measure of inflation to come into line with the rest of Europe and many other countries around the world in 2003. Each month the ONS announce the CPI and the RPI.

Table 21.2 shows that food is not as important in the typical household budget in 2012 as it was in 1987, and similarly tobacco and alcohol are also less important. Housing, though, has become more important, perhaps reflecting the interest people have in their homes (there is a large proportion of the population in the UK who own their own homes unlike the situation in

other parts of Europe where renting is more common). It is also worth noting the increased importance of leisure services, which includes spending on holidays, which have risen significantly in importance since 1987.

Some goods and services that are not as representative any more are removed from the basket. This is measured by the amount spent on these goods and services. If consumer expenditure on items exceeds £400 million annually then these are normally included in the basket but if expenditure falls below £100 million then there has to be a very good reason put forward to keep these goods in the basket. In recent years this has included microwave ovens which first made an appearance in the index in the 1980s. They are still popular, say the ONS, but their reliability and the falling prices means that the amount spent on them has

TABLE 21.2 Product Group Weightings since 1987

| RPI GROUP | 1987 | 1992 | 1997 | 2002 | 2007 | 2011 | 2012 |
|-------------------------------------|------|------|------|------|------|------|------|
| Food | 167 | 152 | 136 | 114 | 105 | 118 | 114 |
| Catering | 46 | 47 | 49 | 52 | 47 | 47 | 47 |
| Alcohol | 76 | 80 | 80 | 68 | 66 | 60 | 56 |
| Tobacco | 38 | 36 | 34 | 31 | 29 | 28 | 29 |
| Housing | 157 | 172 | 186 | 199 | 238 | 238 | 237 |
| Fuel and light | 61 | 47 | 41 | 31 | 39 | 42 | 46 |
| Household goods | 73 | 77 | 72 | 73 | 66 | 65 | 62 |
| Household services | 44 | 48 | 52 | 60 | 65 | 63 | 67 |
| Clothing and footwear | 74 | 59 | 56 | 51 | 44 | 44 | 45 |
| Personal goods and services | 38 | 40 | 40 | 43 | 39 | 38 | 39 |
| Motoring expenditure | 127 | 143 | 128 | 141 | 133 | 137 | 131 |
| Fares and other travel costs | 22 | 20 | 20 | 20 | 20 | 20 | 23 |
| Leisure goods | 47 | 47 | 47 | 48 | 41 | 36 | 33 |
| Leisure services | 30 | 32 | 59 | 69 | 68 | 64 | 71 |

Weights are specified as parts per 1,000 of the all items' RPI

reduced and as such they have been removed from the basket. Other items that have dropped out include film for 35 mm cameras (the cameras themselves dropped out in 2007), top 40 CD singles and television repair, as technology changes the way we live. The ONS says that people prefer to download music rather than buy singles and households tend to replace TVs rather than get them repaired.

The goods and services that have come into the basket are a

reflection of the growing 'café culture' in the UK. Muffins have been brought into the basket to represent croissants and cakes that tend to be bought with coffee in cafés. The ONS has also reported that the UK tends to be more health conscious; fruit smoothies are now more popular so they enter the basket as do small fruits such as satsumas and clementines. Perhaps not so healthy are crates of 20 bottles of lager. However, they are now in the basket. The ONS says that the once

popular 'stubbie' is now less so and people are tending to buy lager in different ways. Hot rotisserie cooked chickens, Blu-ray discs, bottled rosé wine, MP4 players, baby wipes, television and internet access subscriptions, tablet computers and soft continental cheeses have also made it into the basket but single cream, candy coated chocolates, imported lamb loin chops, glass ovenware casserole dishes, step ladders and wine boxes have been removed.

To take another example, in 2004 the ONS introduced digital cameras into the CPI for the first time. As well as the problems associated with introducing a new good into the index that we discussed earlier, digital cameras are also subject to very rapid technological progress – features such as zoom and the number of megapixels in the pictures taken keep on improving rapidly and the sale of digital cameras has now been impacted by the developments in camera technology which is incorporated in smart phones. Thus, while the average price of a digital camera or a smart phone might remain the same over a period, the average quality may have risen substantially. The ONS attempts to correct for this by a method known as hedonic quality adjustment. This involves working out the average characteristics (e.g. LCD screen size, number of megapixels, zoom features, etc.) of the average digital camera and adjusting the price when one of these average characteristics increases.

Despite these efforts, changes in quality remain a problem, because quality is so hard to measure.

Relevance of the CPI A final problem with the index is that people may not see the reported CPI measure of inflation as relevant to their particular situation. This is because their spending patterns are individual and might not be typical of the representative pattern on which the official figures are based. For example, if an individual spent a high proportion of her income on fuel and her mortgage, the effect of price rises in gas, electricity, petrol and a rise in mortgage rates would have a disproportionate effect on her own experience of inflation. This different perception of inflation can have an effect on expectations, the importance of which we will see in later chapters.

Because of the different perceptions that people have about inflation and how it affects them, the ONS published a Personal Inflation Calculator (PIC) in 2007. The PIC allows users to be able to input their own details such as what their personal monthly expenditure is, how much they spend on food, meals out, alcohol, clothing and footwear, fuel for transport and so on. In addition, the calculator looks at what is spent on utilities such as water, council tax, vehicles, holidays and housing.

The ONS hopes that the PIC will help people to develop more of an interest in how inflation is calculated and be more aware of how the reality of price rises in their lives might differ to the officially published figures.

The CPI, the Harmonized Index of Consumer Prices and the Retail Prices Index

Before the end of 2003, it was more usual in the UK to measure prices using an index known as the retail prices index (RPI). In December 2003, however, the UK Chancellor of the Exchequer announced that all policy announcements concerning inflation and prices would relate to the CPI rather than the RPI, so that movements in the CPI became the main measure of inflation. What is the difference between the CPI and the RPI, and why the switch?

The RPI is a price index constructed in the way described above, and differs from the CPI mainly in the goods and services included in the basket and in the coverage of households. In particular, the CPI excludes a number of items that are included in the RPI, mainly related to housing, such as council tax (a local government tax) and house mortgage interest payments. These items are excluded because if council taxes rise or mortgage payments rise because interest rates have risen, then the inflation rate as measured by the RPI will rise, even though underlying inflationary pressures in the economy may not have changed. Also, the CPI covers all private households, whereas the RPI excludes the top 4 per cent by income and pensioner households who derive at least three-quarters of their income from state benefits. The CPI also includes the residents of institutional households such as student hostels, and also foreign visitors to the UK. This means that it covers some items that are not in the RPI, such as stockbrokers' fees, university accommodation fees and foreign students' university tuition fees. The two indices also differ in some of the very fine details of the way in which prices are measured (such as allowance for quality adjustment).

What were the reasons for introducing the CPI? One reason is that some economists believe that it is closer to the concept of the overall price level employed in macroeconomic analysis – although others would argue that it may be misleading in the way in which it understates housing costs by excluding mortgage interest payments and council taxes. The main reason for its adoption, however, is that the construction of the CPI is identical to that used in the European Union (EU) to construct price indices for

TABLE 21.3**Inflation Rates Across Europe (2005 = 100) October 2013**

| | |
|---|-------------|
| Belgium | 1.4 |
| Bulgaria | 1.0 |
| Czech Republic | 1.6 |
| Denmark | 0.8 |
| Germany (including former GDR from 1991) | 1.7 |
| Estonia | 3.5 |
| Ireland | 0.7 |
| Greece | -0.4 |
| Spain | 2.0 |
| France | 1.1 |
| Italy | 1.6 |
| Cyprus | 0.8 |
| Latvia | 0.3 |
| Lithuania | 1.6 |
| Luxembourg | 1.9 |
| Hungary | 2.5 |
| Malta | 1.4 |
| Netherlands | 2.9 |
| Austria | 2.3 |
| Poland | 1.1 |
| Portugal | 0.8 |
| Romania | 3.7 |
| Slovenia | 2.2 |
| Slovakia | 2.0 |
| Finland | 2.5 |
| Sweden | 0.5 |
| United Kingdom | 2.7 |
| Iceland | 4.5 |
| Norway | 1.8 |
| Switzerland | 0.0 |
| Croatia | 2.9 |
| Turkey | 7.3 |

other EU member countries. Because these price indices are harmonized in their construction across countries, they are known in Europe as harmonized indices of consumer prices, or HICPs. This allows direct comparison of inflation rates across EU member states. Such comparisons are not possible using national consumer price indices due to differences in index coverage and construction. The CPI published by the UK Office for National Statistics is in fact the UK HICP.

In Table 21.3 we have listed the HICP as of October 2013 for each of the members of the European Union, and for Switzerland and Turkey. Because we know that the price index has been constructed using the same conventions for each country, we can compare the figures directly.

The GDP Deflator versus the Consumer Prices Index

In the preceding chapter we examined another measure of the overall level of prices in the economy – the GDP deflator. The GDP deflator is the ratio of nominal GDP to real GDP. Because nominal GDP is current output valued at current prices and real GDP is current output valued at base-year prices, the GDP deflator reflects the current level of prices relative to the level of prices in the base year.

Economists and policymakers monitor both the GDP deflator and the CPI to gauge how quickly prices are rising. Usually, these two statistics tell a similar story. Yet there are two important differences that can cause them to diverge.

The first difference is that the GDP deflator reflects the prices of all goods and services *produced domestically*, whereas the CPI reflects the prices of all goods and services *bought by consumers*. For example, suppose that the price of an aeroplane produced by Dassault, a French aerospace firm and sold to the French Air Force rises. Even though the aeroplane is part of GDP in France, it is not part of the basket of goods and services bought by a typical consumer. Thus, the price increase shows up in the GDP deflator for France but not in France's CPI.

As another example, suppose that Volvo raises the price of its cars. Because Volvos are made in Sweden, the car is not part of French GDP. But French consumers buy Volvos, and so the car is part of the typical consumer's basket of goods. Hence, a price increase in an imported consumption good, such as a Volvo, shows up in the CPI but not in the GDP deflator.

This first difference between the CPI and the GDP deflator is particularly important when the price of oil changes. Although the United Kingdom does produce some oil, as with all of Europe and also North America, much of the oil used in the UK is imported from the Middle East. As a result, oil and oil products such as petrol and heating oil comprise a much larger share of consumer spending than they do of GDP. When the price of oil rises, the CPI rises by much more than does the GDP deflator.

The second and subtler difference between the GDP deflator and the CPI concerns how various prices are weighted to yield a single number for the overall level of prices. The CPI compares the price of a *fixed* basket of goods and services with the price of the basket in the base year. Whilst, as we have seen, the ONS revise the basket of goods on a regular basis, in contrast, the GDP deflator compares the price of *currently produced* goods and services with the price of the same goods and services in the base year. Thus, the group of goods and services used to compute the GDP deflator changes automatically over time. This difference is not important when all prices are changing proportionately. But if the prices of different goods and services are changing by varying amounts, the way we weight the various prices matters for the overall inflation rate.

Figure 21.1 shows the UK inflation rate as measured by both the GDP deflator and the CPI for each year from 1996 until 2012. You can see that whilst there are some occasions when the two diverge they do tend to move together. Figure 21.2 shows the GDP deflator for the EU 27 from 2010 to 2012 as a means of comparison.

SELF TEST Explain briefly what the consumer prices index is trying to measure and how it is constructed.

FIGURE 21.1

Two Measures of Inflation

This figure shows UK inflation as measured by the GDP deflator and the CPI using annual data since 1996. Notice that the two measures of inflation generally move together.

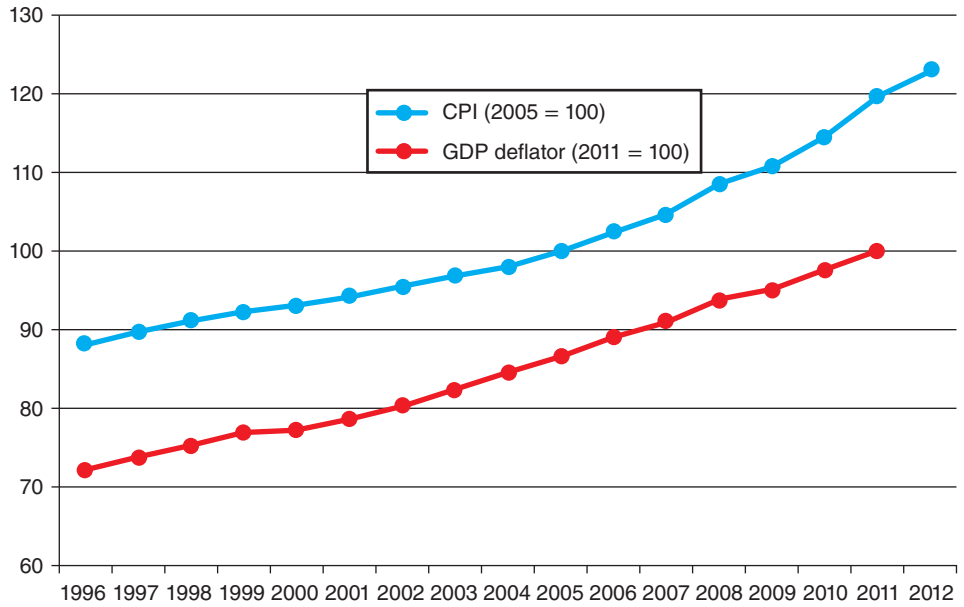
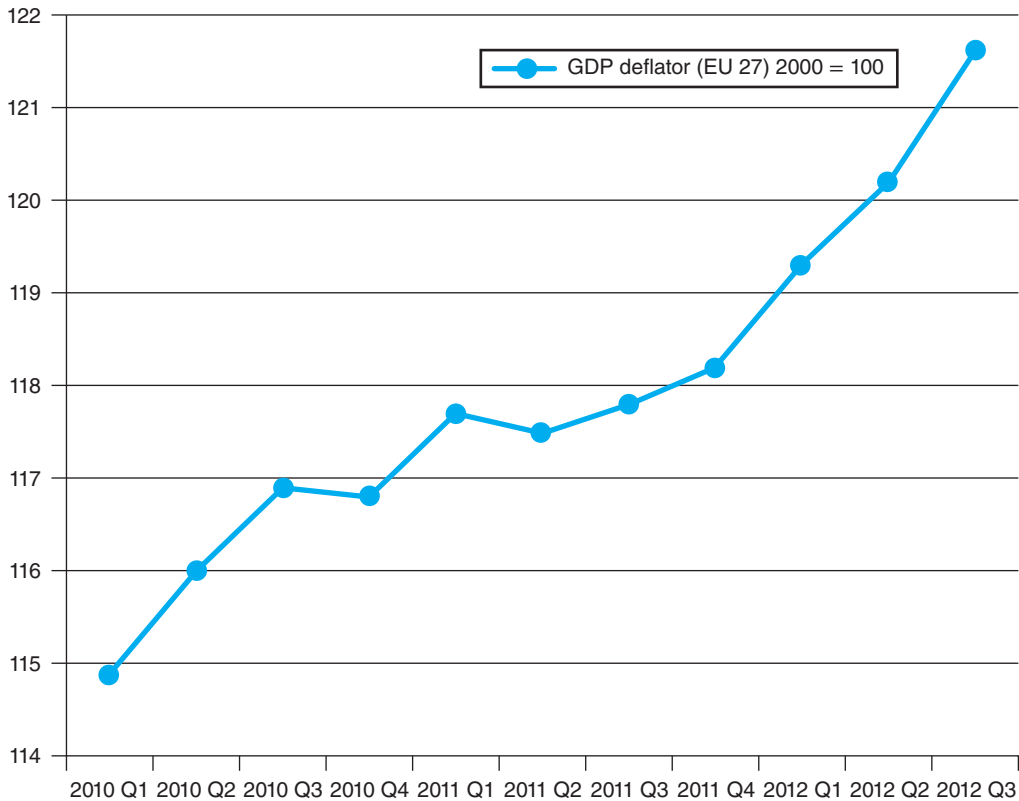


FIGURE 21.2

GDP Deflator (EU 27)

This figure shows the GDP deflator for the EU 27 countries by quarter between 2010 and 2012.



CORRECTING ECONOMIC VARIABLES FOR THE EFFECTS OF INFLATION

The purpose of measuring the overall level of prices in the economy is to permit comparisons of monetary figures from different points in time. Now that we know how price indices are calculated, let's see how we might use such an index to compare a certain figure from the past to a figure in the present.

Money Figures from Different Times

We first return to the issue of MPs' salary. Was the salary of £400 a year, first paid to MPs in 1911, high or low compared to the salaries of today's MPs?

To answer this question, we need to know the level of prices in 1911 and the level of prices today. Part of the increase in salaries merely compensates MPs for the higher level of prices today. To compare the salary to those of today's MPs, we need to inflate the 1911 salary to turn 1911 pounds into today's pounds. A price index determines the size of this inflation correction.

The formula for turning pound figures from *year T* into today's pounds is the following (assuming the price level today and the price level in *year T* are measured against the same base year):

$$\text{Amount in today's pounds} = \text{Amount in year } T \text{ pounds} \times \frac{\text{Price level today}}{\text{Price level in year } T}$$

A price index such as the CPI measures the price level and determines the size of the inflation correction. Let's apply this formula to MPs' wages. Statistics from the ONS show a price index of 9.6 for 1911 and 973.6 for December 2012. Thus the overall level of prices has risen by a factor of 101.4 (which equals 973.6/9.6). We can use these numbers to measure MP's salary in 2012 pounds, as follows:

$$\begin{aligned} \text{Salary in December 2012 pounds} &= \text{Salary in 1911 pounds} \times \frac{\text{Price level in December 2012}}{\text{Price level in 1911}} \\ &= £400 \times \left(\frac{973.6}{9.6} \right) \\ &= £40,567 \end{aligned}$$

We find that the salary paid to an MP in 1911 is equivalent to a salary today of £40,567. This is more than the median annual salary across all occupations in April 2012 of £26,500 and is nearly two-thirds of what MPs were paid in 2012: £65,738. In that sense, MPs have done well.

Indexation

As we have just seen, price indices are used to correct for the effects of inflation when comparing monetary figures from different times. This type of correction shows up in many places in the economy. When some money amount is automatically corrected for inflation by law or contract, the amount is said to be **indexed** for inflation.

indexed the automatic correction of a money amount for the effects of inflation by law or contract

For example, many long-term contracts between firms and unions include partial or complete indexation of the wage to the CPI. Such a provision is called a *cost-of-living allowance*, or COLA. A COLA automatically raises the wage each year based on the CPI or other measure such as the RPI, at a particular point in time.

Income tax brackets – the income levels at which the tax rates change – are also often moved annually in line with inflation, although, in most countries, they are not formally indexed. Indeed, there are many

CASE STUDY Adjusting for Inflation: Use the Force

What was the most popular film of all time? The answer might surprise you. Film popularity is usually gauged by worldwide box office receipts. By that measure, as of the end of 2012, *Avatar* (released in 2009) was the number one film of all time grossing \$2.78 billion, followed by *Titanic* (1993 – \$2.19 billion); *Marvel's the Avengers* (2012 – \$1.5 billion); *Harry Potter and the Deathly Hallows, Part 2* (2011 – \$1.33 billion); *Lord of the Rings: Return of the King* (2003 – \$1.19 billion); *The Dark Knight Rises* (2012 – \$1.08 billion); and *Skyfall* (2012 – \$1.08 billion). But this ranking ignores an obvious but important fact: prices, including those of cinema tickets, have been rising over time. When we correct box office receipts for the effects of inflation, the story is very different.

Table 21.4 shows the top ten films of all time ranked by inflation-adjusted worldwide box office receipts in US dollars. The original *Star Wars*, released in 1977, moves up to number two, while *The Sound of Music* takes the number three spot and the number one film of all time is the American Civil War drama, *Gone with the Wind* (\$1.64 billion adjusted) which was released in 1939 and is well ahead of *Avatar* (\$533 million adjusted), which doesn't even make the top ten (appearing at number 14). In the 1930s, before everyone had televisions in their homes, cinema attendance was about three or four times what it is today. But the films from that era rarely show up in popularity rankings because ticket prices were only a fraction of what they are today. Scarlett and Rhett, the main characters in *Gone with the Wind*, fare a lot better once we correct for the effects of inflation, as do Mowgli, Balou and King Louie of *The Jungle Book* fame (number 9 in the inflation-adjusted list). However, television ownership (indeed multiple television ownership) was widespread by 1977 when *Star Wars* was released, and so the ranking of this film as number two in the inflation-adjusted table perhaps makes Luke Skywalker an even more impressive figure.

The number one film of all time is the American Civil War drama, Gone with the Wind (\$1.64 billion adjusted) which was released in 1939.



TABLE 21.4

The Most Popular Films of All Time, Inflation-Adjusted

| Film | Year of release |
|--|-----------------|
| 1. <i>Gone with the Wind</i> | 1939 |
| 2. <i>Star Wars</i> | 1977 |
| 3. <i>The Sound of Music</i> | 1965 |
| 4. <i>E.T.</i> | 1982 |
| 5. <i>Titanic</i> | 1997 |
| 6. <i>The Ten Commandments</i> | 1956 |
| 7. <i>Jaws</i> | 1975 |
| 8. <i>Doctor Zhivago</i> | 1965 |
| 9. <i>The Jungle Book</i> | 1967 |
| 10. <i>Snow White and the Seven Dwarfs</i> | 1937 |

ways in which the tax system is not indexed for inflation, even when perhaps it should be. We discuss these issues more fully when we discuss the costs of inflation later in this book.

Real and Nominal Interest Rates

Correcting economic variables for the effects of inflation is particularly important, and somewhat tricky, when we look at data on interest rates. When you deposit your savings in a bank account, you will earn interest on your deposit. Conversely, when you borrow from a bank in order to buy a car, you will pay interest on the loan. Interest represents a payment in the future for a transfer of money in the past. As a result, interest rates always involve comparing amounts of money at different points in time. To fully understand interest rates, we need to know how to correct for the effects of inflation.

Let's consider an example. Suppose that Carla deposits €1,000 in a bank account that pays an annual interest rate of 10 per cent. After a year passes, Carla has accumulated €100 in interest. Carla then withdraws her €1,100. Is Carla €100 richer than she was when she made the deposit a year earlier?

The answer depends on what we mean by 'richer'. Carla does have €100 more than she had before. In other words, the number of euros has risen by 10 per cent. But if prices have risen at the same time, each euro now buys less than it did a year ago. Thus, her purchasing power has not risen by 10 per cent. If the inflation rate was 4 per cent, then the amount of goods she can buy has increased by only 6 per cent. And if the inflation rate was 15 per cent, then the price of goods has increased proportionately more than the number of euros in her account. In that case, Carla's purchasing power has actually fallen by 5 per cent.

The interest rate that the bank pays is called the **nominal interest rate**, and the interest rate corrected for inflation is called the **real interest rate**. We can write the relationship among the nominal interest rate, the real interest rate and inflation as follows:

nominal interest rate the interest rate as usually reported without a correction for the effects of inflation
real interest rate the interest rate corrected for the effects of inflation

$$\text{Real interest rate} = \text{Nominal interest rate} - \text{Inflation rate}$$

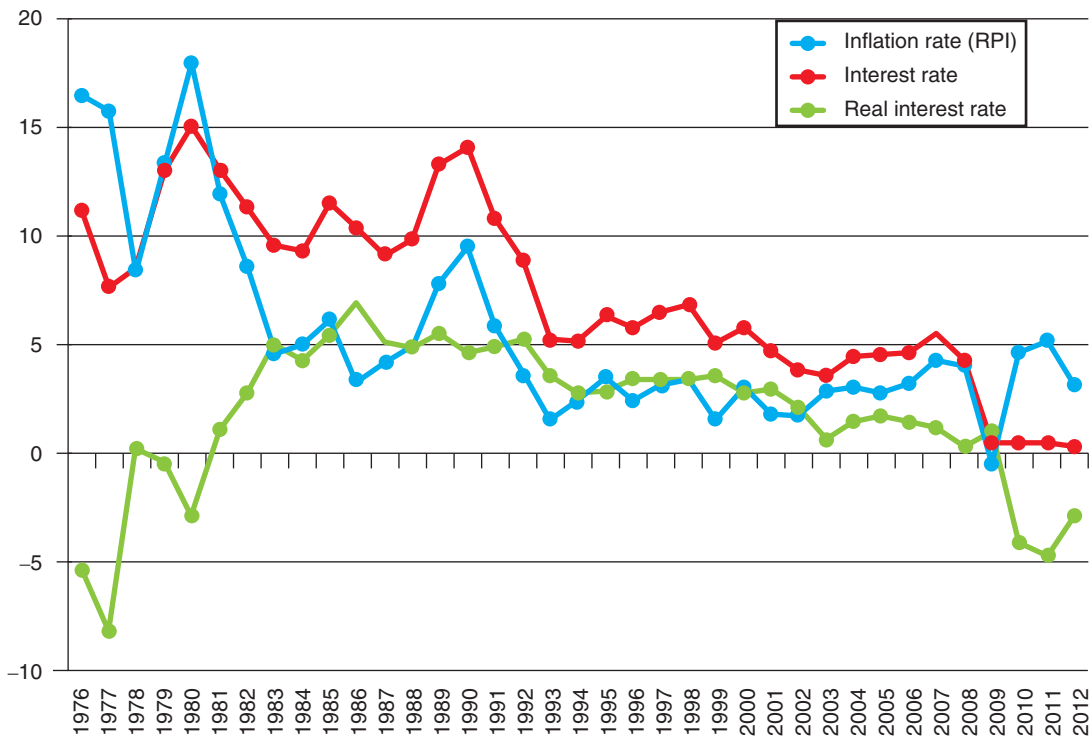
The real interest rate is the difference between the nominal interest rate and the rate of inflation. The nominal interest rate tells you how fast the number of pounds or euros in your bank account rises over time. The real interest rate tells you how fast the purchasing power of your bank account rises over time.

Figure 21.3 shows the nominal interest rate (measured as the average rate of discount on three-month Treasury bills), the inflation rate (measured by the RPI) and the real interest rate (the difference between the two). You can see that real and nominal interest rates do not always move together. For example, in the late 1970s, nominal interest rates were high, but because inflation was higher, real interest rates were low. It is noticeable that since the financial crisis in 2008, real interest rates have been negative with inflation eroding people's savings more quickly than nominal interest payments increased them. By contrast, in the 1990s, nominal interest rates were low but because inflation was also low, real interest rates were relatively high. In the coming chapters, when we study the causes and effects of changes in interest rates, it will be important for us to keep in mind the distinction between real and nominal interest rates.

SELF TEST Why is it important to adjust for inflation when considering wage increases and the returns from any financial investment?

FIGURE 21.3**Real and Nominal Interest Rates**

This figure shows the inflation rate (RPI), nominal and real interest rates using average annual changes since 1976. The nominal interest rate is the rate on a three-month Treasury bill. The real interest rate is the nominal interest rate minus the annual inflation rate as measured by the retail prices index. Notice that nominal and real interest rates do not always move together.



CONCLUSION

Throughout recent history, the real values behind the currencies of most major industrialized countries, such as the pound, euro and dollar, have not been stable. Persistent increases in the overall level of prices in advanced economies such as those of the UK, the other countries of Europe and of the USA have been the norm. Such inflation reduces the purchasing power of each unit of money over time. When comparing figures in monetary or nominal terms from different times, it is important to keep in mind that a pound or euro today is not the same as a pound or euro ten years ago or, most likely, ten years from now.

This chapter has discussed how economists measure the overall level of prices in the economy and how they use price indices to correct economic variables for the effects of inflation. This analysis is only a starting point. We have not yet examined the causes and effects of inflation or how inflation interacts with other economic variables. To do that, we need to go beyond issues of measurement. Indeed, that is our next task. Having explained how economists measure macroeconomic quantities and prices in the past two chapters, we are now ready to develop the models that explain long-run and short-run movements in these variables.

IN THE NEWS



Accounting for Quality Change

Economists and statisticians are always looking to find ways of improving data, which is so crucial to decision making and policy. Accurate and reliable inflation data are one of the areas that have been subject to considerable review. In particular the issue of how changes in quality affect measures of inflation has been at the forefront of debate for a number of years. This article looks at the use of hedonics in the calculation of price indices.

Hedonics and Price Indices

We all know that prices change over time. The reason for price changes can be varied but some price changes may reflect improvements in the quality of a product as technology advances ever more rapidly. Does this mean that a TV bought in 2012 for €600 represents better value for money (and therefore a lower price) than one priced at €450 in 2009? Taking into account this issue is the subject of hedonics. Hedonics is a statistical device to account for the changing quality of products when calculating price movements. It has stirred some debate on both sides of the Atlantic. In the United States for example, the *Wall Street Journal* reported on the case of 27-inch TV sets. Tim LaFleur, a commodity specialist for televisions working with the US Bureau of Labor Statistics, the body which compiles CPI data, estimated that the price of a TV set with a nominal price of \$329.99 had improved in terms of the quality of the screen, among other things, which increased its value by \$135. When running the data through a computer program, LaFleur concluded that the price of the TV had actually fallen by 29 per cent.

There are other aspects to the issue as well. The increasing concern with the effects of humans on the planet means that there is a

drive to improve energy efficiency. Developments in technology are also helping to improve health and safety, the improvements in safety records of cars being one example of this. If manufacturers produce a washing machine that improves the efficiency of energy use by 10 per cent, for example, then even if the nominal price of the product rose, the use of hedonics might show that the price of washing machines did not rise at all. Equally, if a car is now safer to use and protects the driver and passengers, then the value of this is far greater than any increase in price to cover the cost of fitting such features. The adjusted price of the car, therefore, falls.

The ONS has been using hedonics as part of its calculation of price indices since 2004. One news report at the time suggested that the use of hedonics could save the Chancellor around £100 million a year, as a result of having to pay less to savers and pensioners. Initially the technique was applied to the RPI relating to digital cameras and computers. It led to a cut in the RPI by around 0.05 per cent, according to the ONS, but this equates to a saving for the government of £50 million in payments to those on index linked benefits, £45 million on interest payments for index linked gilts and further savings on tax allowances which may not be uprated. Whilst the initial

introduction of hedonics at the ONS caused a relatively small change in the RPI, there have been suggestions that a wider application of the technique across more goods could reduce the annual rate of inflation by between 0.25 and 0.5 per cent.

There have been concerns expressed, therefore, that the use of this technique presents a distorted picture of how price changes affect the economy. A report published in the US in 1996 suggested that underestimation of the improvements in quality might mean that price indices were overstated by as much as 0.6 per cent per year. There are some concerns, however, that the technique could be exploited for political purposes; whilst hedonics has a firm root in statistical methodology there is some judgement applied about the nature of the quality improvements and there are concerns that these might be used to artificially lower published inflation rates. Given that inflation figures are used to index link certain payments from government and as the basis for wage claims, any depression in the index might be viewed with some suspicion.

Proponents of the use of hedonics argue that far from distorting the picture, the use of the technique gives a far more accurate picture of what is going on in the economy. They point out that bodies such as

(Continued)

the Royal Statistical Society make observations on the way the ONS gathers and interprets data and that it is an important way of measuring quality changes in relation to prices. Technology is improving so rapidly that a failure to do so would over inflate prices. They point to the way in which products like digital cameras, mobile phones, TVs and computers have changed in recent years. In many cases, improvements in technology also lead to a reduction in price which will affect price indices. The question is should these price indices be even lower? Take tablet

computers, which at the time of writing have the computing power to run advanced, high-quality graphics games, have massive memory, roaming internet access, extensive music and video files, cameras and video recorders and players, and are priced at around £500 in comparison to bulky, low graphics, slow machines that basically ran word-processing and other office applications but which were priced over £1,000 five years ago. Surely the improvements in speed, quality, capacity and flexibility are worth more to consumers than the change in price of £500?

Ultimately the debate comes down to the accuracy of the judgements made by statisticians in being able to accurately account for the improvements in quality of products. If these judgements are accurate and widely accepted then incorporating hedonics into the index could improve its accuracy. If, however, such judgements are subject to political interference then some sections of society could be made worse off.

Questions

- 1 **Why is it important to take changes in quality into account when measuring price changes over time?**
- 2 **What sort of quality improvements do you think statisticians ought to account for in generating more accurate measures of price changes over time?**
- 3 **Why is index linking important to people whose incomes are 'fixed' (i.e. not subject to increases linked to productivity improvements)?**
- 4 **Why might the incorporation of hedonics into measurement of prices over time affect those whose incomes are index linked?**
- 5 **Do you think that incorporating hedonics into measurement of price changes is a good thing or too likely to be subject to political interference?**

Technology changes over time as do prices but how can we take into account changes in quality over time? Hedonics may provide the answer.



SUMMARY

- The CPI shows the changes in the prices of a basket of goods and services relative to the prices of the same basket in the base year. The index is used to measure the overall level of prices in the economy. The percentage change in the CPI measures the inflation rate.
- The CPI is an imperfect measure of the cost of living for three reasons. First, it does not take into account consumers' ability to substitute towards goods that become relatively cheaper over time. Second, it does not take into account increases in the purchasing power of money due to the introduction of new goods. Third, it is distorted by unmeasured changes in the quality of goods and services. Because of these measurement problems, the CPI overstates true inflation.
- Although the GDP deflator also measures the overall level of prices in the economy, it differs from the CPI because it includes goods and services produced rather than goods and services consumed. As a result,

imported goods affect the CPI but not the GDP deflator. In addition, while the CPI uses a fixed basket of goods, the GDP deflator automatically changes the group of goods and services over time as the composition of GDP changes.

- Money figures (e.g. in euros) from different points in time do not represent a valid comparison of purchasing power. To compare a money figure from the past to a money figure today, the older figure should be inflated using a price index.
- Various laws and private contracts use price indices to correct for the effects of inflation.
- A correction for inflation is especially important when looking at data on interest rates. The nominal interest rate is the interest rate usually reported; it is the rate at which the amount of money in a savings account increases over time. In contrast, the real interest rate takes into account changes in the value of the money over time. The real interest rate equals the nominal interest rate minus the rate of inflation.

QUESTIONS FOR REVIEW

- 1 What are the five stages of constructing a price index?
- 2 Why do statisticians use weighting in constructing price indices?
- 3 Which do you think has a greater effect on the consumer prices index: a 10 per cent increase in the price of chicken or a 10 per cent increase in the price of caviar? Why?
- 4 Describe the three problems that make the CPI an imperfect measure of the cost of living.
- 5 Why do statisticians change the composition of the basket of goods used in constructing a price index from time to time?
- 6 Assume that the price of a bottle of wine in 1990 was €3.50 and in 2012 it is €8.50. Further assume that the price index in 1990 was 95 and in 2012 was 160. Was wine cheaper in 1990 than in 2012? Explain.
- 7 If the price of a French Navy submarine rises, is the French consumer prices index or the French GDP deflator affected more? Why?
- 8 Over a long period of time, the price of a box of chocolates rose from €1 to €6. Over the same period, the CPI rose from 150 to 300. Adjusted for overall inflation, how much did the price of the box of chocolates change?
- 9 Explain the meaning of nominal interest rate and real interest rate. How are they related?
- 10 Why is knowledge of real interest rates of importance to people who rely on their savings for their income?

PROBLEMS AND APPLICATIONS

- 1 Suppose that people consume only three goods, as shown in this table:

| | Tennis balls | Tennis racquets | Cola |
|----------------|--------------|-----------------|------|
| 2014 price (€) | 2 | 40 | 1 |
| 2014 quantity | 100 | 10 | 200 |
| 2015 price (€) | 2 | 60 | 2 |
| 2015 quantity | 100 | 10 | 200 |

- a. What is the percentage change in the price of each of the three goods? What is the percentage change in the overall price level?
 - b. Do tennis racquets become more or less expensive relative to cola? Does the well-being of some people change relative to the well-being of others? Explain.
- 2 Suppose that the residents of Vegopia spend all of their income on cauliflower, broccoli and carrots. In 2014 they buy 100 heads of cauliflower for €200, 50 bunches of broccoli for €75 and 500 carrots for €50. In 2015 they buy 75 heads of cauliflower for €225, 80 bunches of broccoli for €120 and 500 carrots for €100. If the base year is 2014, what is the CPI in both years? What is the inflation rate in 2015?
 - 3 Go to the website of the UK Office for National Statistics (<http://www.statistics.gov.uk>) or Eurostat (<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>) and find data on the CPI. By how much has the index including all items risen over the past year for your country? For which categories of spending have prices risen the most? The least? Have any categories experienced price declines? Can you explain any of these facts?
 - 4 Which of the problems in the construction of the CPI might be illustrated by each of the following situations? Explain.
 - a. the increase in downloading movies
 - b. the introduction of air bags in cars
 - c. increased personal computer purchases in response to a decline in their price
 - d. increased use of digital cameras in smart phones
 - e. greater use of fuel-efficient cars after petrol prices increase.
 - 5 Suppose the government were to determine the level of the state retirement pension in the UK so that it increased each year in proportion to the increase in the

CPI (even though most economists believe that the CPI overstates actual inflation).

- a. If the elderly consume the same market basket as other people, would such a policy provide the elderly with an improvement in their standard of living each year? Explain.
 - b. In fact, the elderly consume more health care than younger people, and health care costs tend to rise faster than overall inflation. What would you do to determine whether the elderly are actually better off from year to year?
- 6** How do you think the basket of goods and services you buy differs from the basket bought by the typical household? Do you think you face a higher or lower inflation rate than is indicated by the CPI? Why?
- 7** In some years in some countries, income tax brackets are not increased in line with inflation. Why do you think the government might do this? (Hint: this phenomenon is known as ‘bracket creep’.)
- 8** When deciding how much of their income to save for retirement, should workers consider the real or the nominal interest rate that their savings will earn? Explain.
- 9** Suppose that a borrower and a lender agree on the nominal interest rate to be paid on a loan. Then inflation turns out to be higher than they both expected.
- a. Is the real interest rate on this loan higher or lower than expected?
 - b. Does the lender gain or lose from this unexpectedly high inflation? Does the borrower gain or lose?
- 10** Explain why nominal and real interest rates might vary over time.

PART 11

THE REAL ECONOMY IN THE LONG RUN

22 PRODUCTION AND GROWTH

When you travel around the world, you see tremendous variation in the standard of living. The average person in a rich country, such as the countries of Western Europe, has an income more than ten times as high as the average person in a poor country, such as India, Indonesia or Nigeria. These large differences in income are reflected in large differences in the quality of life. Richer countries have more cars, more smart phones, more televisions, better nutrition, safer housing, better health care and longer life expectancy.

Even within a country, there are large changes in the standard of living over time. In the United Kingdom over the past century or so, average income as measured by real GDP per person has grown by about 1.3 per cent per year. Although 1.3 per cent might seem small, this rate of growth implies that average income doubles about every 50 years. Because of this growth, average UK income today is about four times as high as average income a century ago. As a result, the typical Brit enjoys much greater economic prosperity than did his or her parents, grandparents and great-grandparents. The story is similar for many other European countries. Figure 22.1 shows GDP per person in the EU 27 since 2003. The trend has been upwards until the recession struck but despite this the average person in the EU 27 is still better off in 2011 than they were in 2003.

Growth rates vary substantially from country to country. In some East Asian countries, such as Singapore, South Korea and Taiwan, average income has risen about 7 per cent per year in recent decades. At this rate, average income doubles about every ten years. These countries have, in the length of one generation, gone from being among the poorest in the world to being among the richest. In contrast, in some African countries, such as Chad, Ethiopia and Nigeria, average income has been stagnant for many years.

What explains these diverse experiences? How can the rich countries be sure to maintain their high standard of living? What policies should the poor countries pursue to promote more rapid growth in order to join the developed world? These are among the most important questions in macroeconomics. As the Nobel Prize winning economist, Robert Lucas, once put it, 'The consequences for human welfare in questions like these are simply staggering: once one starts to think about them, it is hard to think about anything else.'

In the previous two chapters we discussed how economists measure macroeconomic quantities and prices. In this chapter we start studying the forces that determine these variables. As we have seen, an economy's gross domestic product (GDP) measures both the total income earned in the economy and the

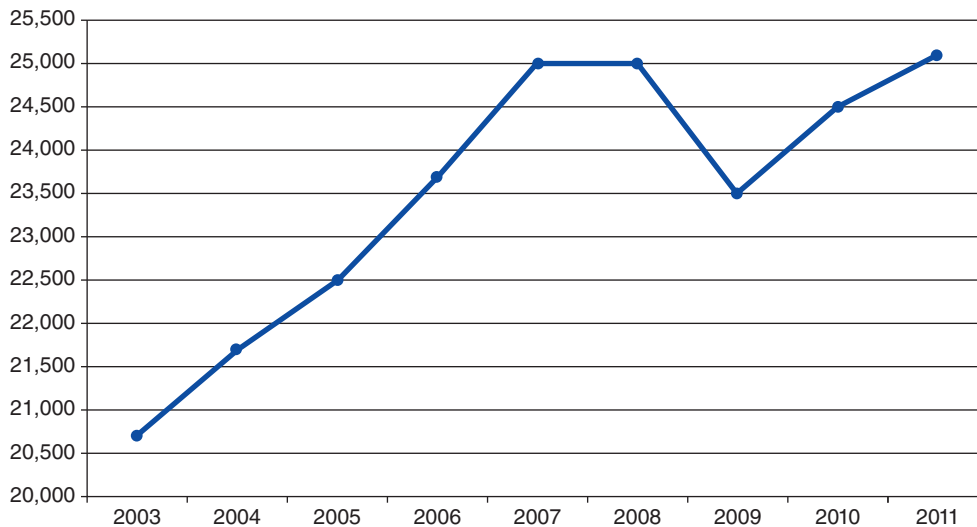
total expenditure on the economy's output of goods and services. The level of real GDP is a good gauge of economic prosperity, and the growth of real GDP is a good gauge of economic progress. Here we focus on the long-run determinants of the level and growth of real GDP. Later in this book we study the short-run fluctuations of real GDP around its long-run trend.

We proceed here in three steps. First, we examine international data on real GDP per person. These data will give you some sense of how much the level and growth of living standards vary around the world. Second, we examine the role of *productivity* – the amount of goods and services produced for each hour of a worker's time. In particular, we see that a nation's standard of living is determined by the productivity of its workers, and we consider the factors that determine a nation's productivity. Third, we consider the link between productivity and the economic policies that a nation pursues.

FIGURE 22.1

EU 27 GDP at Market Prices, Purchasing Power Standard Per Capita, 2003–2011 (€)

The figure shows GDP per person for the 27 EU countries over the period 2003–2011. Average incomes rose up to 2009 after which the recession led to a fall in GDP per capita.



ECONOMIC GROWTH AROUND THE WORLD

As a starting point for our study of long-run growth, let's look at the experiences of some of the world's economies. Table 22.1 shows real GDP per person in 31 selected countries.

The data in the table show that living standards vary widely from country to country and have changed at very different rates over the last 50 years. Income per person in the United States, for instance, is about 32 times that in India even though India's per capita income has increased by almost 2,405 per cent since 1961. The poorest countries have very low average levels of income; the per capita income in the Democratic Republic of Congo has only increased by 18 per cent since 1961. Japan is one of the countries that has seen some of the most significant increases in per capita income rising by over 8,000 per cent since 1961 but Norway, Ireland and Brazil have also seen notable increases in per capita income which have been over 6,000 per cent. In 1961 Japan was not a rich country. Japan's average income is lower than Greece and not that much higher than Mexico's. Data from Barro and Xavier Sala-i-Martin published in 1995 (Robert J. Barro and Xavier Sala-i-Martin (1995). *Economic Growth*. New York: McGraw-Hill) suggested that

Japan's income in 1890 was less than in India today. But because of its spectacular growth, Japan is now an economic superpower, with average income more than 4.5 times that of Mexico, almost 1.8 times that of Greece and similar to that of France and Germany. The typical resident of many African countries continues to live in abject poverty, with average incomes in countries like Burkina Faso, Burundi and the Central African Republic at very low levels.

TABLE 22.1 Real GDP Per Capita, Current US\$, Selected Countries

| Country Name | 1961 | 1971 | 1981 | 1991 | 2001 | 2011 | % change |
|--------------------------|-------|-------|--------|--------|--------|---------|----------|
| Algeria | 221 | 358 | 2,281 | 1,763 | 1,781 | 5,244 | 2,271 |
| Australia | 1,874 | 3,490 | 11,850 | 18,873 | 19,541 | 60,979 | 3,153 |
| Bangladesh | 93 | 129 | 238 | 287 | 356 | 743 | 695 |
| Barbados | 403 | 851 | 3,792 | 6,480 | 9,524 | 13,453 | 3,242 |
| Bolivia | 179 | 254 | 1,077 | 784 | 960 | 2,374 | 1,229 |
| Brazil | 203 | 500 | 2,115 | 2,677 | 3,130 | 12,594 | 6,097 |
| Burkina Faso | 71 | 81 | 240 | 327 | 222 | 600 | 750 |
| Burundi | 68 | 71 | 228 | 204 | 127 | 271 | 300 |
| Central African Republic | 81 | 108 | 297 | 467 | 247 | 489 | 507 |
| Congo, Dem. Rep. | 196 | 268 | 451 | 240 | 92 | 231 | 18 |
| Congo, Rep. | 146 | 234 | 1,076 | 1,110 | 870 | 3,485 | 2,289 |
| Egypt, Arab Rep. | 151 | 225 | 509 | 638 | 1,417 | 2,781 | 1,738 |
| Finland | 1,327 | 2,684 | 10,790 | 24,991 | 24,025 | 48,823 | 3,578 |
| France | 1,446 | 3,126 | 10,917 | 21,270 | 21,812 | 42,377 | 2,830 |
| Germany | – | 3,089 | 9,879 | 22,604 | 22,840 | 44,060 | 1,326 |
| Ghana | 187 | 271 | 375 | 434 | 271 | 1,570 | 739 |
| Greece | 597 | 1,576 | 5,130 | 9,776 | 11,858 | 25,622 | 4,190 |
| Iceland | 1,418 | 3,204 | 14,913 | 26,406 | 27,803 | 43,969 | 3,001 |
| India | 59 | 120 | 275 | 308 | 460 | 1,489 | 2,405 |
| Ireland | 739 | 1,652 | 5,804 | 13,659 | 27,340 | 48,423 | 6,450 |
| Japan | 564 | 2,234 | 10,212 | 28,541 | 32,716 | 45,903 | 8,045 |
| Luxembourg | 2,222 | 4,397 | 13,718 | 35,439 | 45,743 | 114,508 | 5,053 |
| Mexico | 357 | 734 | 3,556 | 3,660 | 6,139 | 10,047 | 2,717 |
| Netherlands | 1,159 | 3,130 | 10,671 | 20,131 | 24,969 | 50,076 | 4,219 |
| Norway | 1,560 | 3,709 | 15,335 | 28,077 | 37,867 | 98,102 | 6,187 |
| Pakistan | 86 | 174 | 337 | 395 | 490 | 1,189 | 1,280 |
| Saudi Arabia | – | 1,127 | 17,617 | 7,879 | 8,849 | 20,540 | 1,722 |
| United Kingdom | 1,453 | 2,526 | 9,142 | 18,387 | 24,836 | 39,038 | 2,588 |
| United States | 2,935 | 5,360 | 13,526 | 23,443 | 35,912 | 48,112 | 1,539 |
| Zambia | 218 | 386 | 672 | 419 | 350 | 1,425 | 555 |
| Zimbabwe | 283 | 405 | 1,058 | 804 | 539 | 757 | 168 |

Because of differences in growth rates, the ranking of countries by income changes substantially over time. As we have seen, Japan is a country that has risen relative to others. One country that has fallen behind is the United Kingdom. In 1870, the United Kingdom was the richest country in the world, with average income about 20 per cent higher than that of the United States and more than twice that of Canada. Today, average income in the United Kingdom is 20 per cent below that of the United States and similar to Canada's. Looking at how the change in per capita income has changed over the past 50 years is one way to look at how countries have grown but it is also usual to look at economic growth from the perspective of the change in GDP over a year. Typical growth rates for developed countries might be between 1 and 2 per cent a year, which may not sound that much but over time can lead to major improvements in standards of living (see the FYI on Compounding). For emerging countries like China, growth rates of between 7 and 12 per cent have been recorded over a period but it has to be remembered that these rates are being recorded from a relatively low base.

These data show that the world's richest countries have no guarantee that they will stay the richest and that the world's poorest countries are not doomed forever to remain in poverty. But what explains these changes over time? Why do some countries zoom ahead while others lag behind? These are precisely the questions that we take up next.

SELF TEST Find out the approximate growth rate of real GDP per person in your country. Find a country that has had faster growth and a country that has had slower growth.

FYI



The Magic of Compounding and the Rule of 70

Suppose you observe that one country has an average growth rate of 1 per cent per year while another has an average growth rate of 3 per cent per year. At first, this might not seem like a big deal. What difference can 2 per cent make?

The answer is, a big difference. Even growth rates that seem small when written in percentage terms seem large after they are compounded for many years.

Consider an example. Suppose that two economics graduates – Milton and Maynard – both take their first jobs at the age of 22 earning €20,000 a year. Milton lives in an economy where all incomes grow at 1 per cent per year, while Maynard

lives in one where incomes grow at 3 per cent per year. Calculations using the principle of compounding show what happens. For Milton, after the first year he earns an additional 1 per cent of €20,000 (€200). In year 2 he earns an additional 1 per cent of €20,200 (€2,402) and so on. To calculate compound figures over time we use the formula $M = P(1 + i)^n$ where M is the total value at the end of the period, P is the initial value, i is the rate of interest (or rate of growth) and n the number of years. After 40 years, therefore, Milton earns $20,000(1 + 0.01)^{40}$ which equals €29,777 a year.

Maynard, on the other hand, earns €20,000 $(1 + 0.03)^{40}$ and earns

€65,240 40 years later. Because of that difference of 2 percentage points in the growth rate, Maynard's salary is more than twice Milton's.

An old rule of thumb, called the *rule of 70*, is helpful in understanding growth rates and the effects of compounding. According to the rule of 70, if some variable grows at a rate of x per cent per year, then that variable doubles in approximately $\frac{70}{x}$ years. In Milton's economy, incomes grow at 1 per cent per year, so it takes about 70 years for incomes to double. In Maynard's economy, incomes grow at 3 per cent per year, so it takes about, $\frac{70}{3}$ or a little over 23 years for incomes to double.

GROWTH THEORY

Over time, economic growth in any country varies. Sometimes the economy experiences strong periods of growth and at other times growth is slower and sometimes the economy shrinks. But over a period of time, a trend can be established which is usually expressed in percentage terms. Trend growth can be found using the following formula:

$$\text{Trend growth} = \frac{\text{GDP}_{t_2} - \text{GDP}_{t_1}}{\text{GDP}_{t_1}} \times 100$$

Which states the trend growth is found by taking GDP in some time period, subtracting GDP from an earlier time period, dividing the result by the value in the initial time period and expressing the figure as a percentage. If we took GDP of the UK for example, in 1860 as time period_{t₁} and then the GDP in 2012 as time period_{t₂}, and did the calculation the answer would give trend growth over that period. Trend growth and actual growth are important factors when we look at business cycles later in the book. Different countries have different trend growth rates. What determines trend growth over time? Why do some countries seem to be able to grow at fast rates and lift their citizens out of poverty whilst others struggle to generate any growth and as a result successive generations of people live in abject poverty? Why is the trend rate in developed countries relatively low?

There have been different theories of economic growth put forward over time, one of the most popular has been that advanced by Robert M. Solow and Trevor Swan in 1956. They identified the rate of human and physical capital and population growth as being key determinants of economic growth. Over the years, economists have looked at other factors that may influence economic growth including the level of macroeconomic stability in an economy (don't worry too much about what this means at this stage, it will become clearer as you read the relevant section of this book), the type of trade policy that exists in a country (is the country outward looking or does it tend to be insular?), the nature and quality of institutions and governance (i.e. how effectively the rule of law operates and how well governments are able to control corruption) in the country concerned, the extent to which violence, war and conflict exist in a country, regional characteristics such as whether the country is part of Europe, North America, Asia or sub-Saharan Africa, geographical factors such as physical resource endowments and climate, the extent to which the country is competitive in international markets, and internal factors such as the amount of productive land available.

You will need to keep in mind that theories of growth are changing as economists carry out more research and have access to more data. The next section will cover the role of productivity in growth which can be referred to as the *neoclassical theory of growth*. Other theories, such as endogenous growth theory, argue that investment in a nation's human capital will be a key driver of economic growth. The reason being that investment in human capital is more likely to lead to increases in technology which in turn help promote efficiency and increases in productivity. This is the main driver of growth rather than trade. The debates surrounding these different theories are beyond the scope of this book but are something that you need to be aware of.

PRODUCTIVITY: ITS ROLE AND DETERMINANTS

Some of the concepts in this chapter revisit those covered in Chapter 6 when we looked at the production function and the concepts of marginal and diminishing marginal product. Your knowledge of these concepts will be useful in this section; the main difference is that we are now thinking on a macroeconomic level rather than a microeconomic one.

Explaining the large variation in living standards around the world is, in one sense, very easy. As we will see, the explanation can be summarized in a single word – *productivity*. But, in another sense, the international variation is deeply puzzling. To explain why incomes are so much higher in some countries than in others, we must look at the many factors that determine a nation's productivity.

Why Productivity is So Important

Let's begin our study of productivity and economic growth by developing a simple model based on an individual who finds himself shipwrecked on a desert island, who we will call Fernando. Because Fernando lives alone, he catches his own fish, grows his own vegetables and makes his own clothes. We can think of Fernando's activities – his production and consumption of fish, vegetables and clothing – as being a simple economy. By examining Fernando's economy, we can learn some lessons that also apply to more complex and realistic economies.

What determines Fernando's standard of living? The answer is obvious. If Fernando is good at catching fish, growing vegetables and making clothes, he lives well. If he is bad at doing these things, he lives poorly. Because Fernando can consume only what he produces, his living standard is tied to his productive ability.

Remember that *productivity* refers to the quantity of goods and services that a worker can produce per specified time period. In the case of Fernando's economy, it is easy to see that productivity is the key determinant of living standards and that growth in productivity is the key determinant of growth in living standards. The more fish Fernando can catch per hour, the more he eats for dinner. If Fernando finds a better place to catch fish, his productivity rises. This increase in productivity makes Fernando better off: he could eat the extra fish, or he could spend less time fishing and devote more time to making other goods he enjoys.

The key role of productivity in determining living standards is as true for nations as it is for Fernando. Recall that an economy's gross domestic product (GDP) measures two things at once: the total income earned by everyone in the economy and the total expenditure on the economy's output of goods and services. The reason why GDP can measure these two things simultaneously is that, for the economy as a whole, they must be equal. Put simply, an economy's income is the economy's output.

Like Fernando, a nation can enjoy a high standard of living only if it can produce a large quantity of goods and services. Western Europeans live better than Nigerians because western European workers are more productive than Nigerian workers. The Japanese have enjoyed more rapid growth in living standards than Mexicans because Japanese workers have experienced more rapidly growing productivity. Indeed, one of the *Ten Principles of Economics* is that a country's standard of living depends on its ability to produce goods and services.

Hence, to understand the large differences in living standards we observe across countries or over time, we must focus on the production of goods and services. But seeing the link between living standards and productivity is only the first step. It leads naturally to the next question: why are some economies so much better at producing goods and services than others?

The Determinants of Economic Growth

Solow and Swan established some key elements of the determinants of economic growth. GDP in any country can be assumed to be an extension of a firm's production function where the level of output is dependent on the factors of production employed. In the analysis that follows we are going to assume just two factors, capital (K) and labour (L) are employed. At a macroeconomic level, therefore, the GDP of a country (Y) is a function of the capital stock (K) and the supply of human capital (L). We will express this as follows:

$$Y = f(K, L)$$

Remember also that the total product curve gets flatter as the amount of each factor input is increased due to the law of diminishing marginal product. The marginal product of labour is the change in output (GDP) which results from the change in labour inputs: MP_L

$$MP_L = \frac{\Delta Y}{\Delta L}$$

And the marginal product of capital is the change in output (GDP) which results from the change in capital inputs:

$$MP_K = \frac{\Delta Y}{\Delta K}$$

The factor endowments that exist in many countries are closely interlinked. We saw in the In the News article in Chapter 19 some criticism of the principle of comparative advantage on the basis that countries with large amounts of cheap labour at their disposal, which should give them a comparative advantage, do not always experience the sort of growth that the theory might suggest. It is not just the amount of labour that is important but the productivity of labour and this is determined by many other factors, not least the amount and quality of capital.

We can do some rearranging of the production function to derive average labour productivity to see how this is dependent on the capital stock. If we take our production function $Y = f(K, L)$ and divide both sides by L we get:

$$\frac{Y}{L} = f\left(\frac{K}{L}, 1\right)$$

$\frac{Y}{L}$ is average labour productivity – the amount of output per worker. Average labour productivity is dependent on the ratio of capital to labour $\frac{K}{L}$. The number 1 is a constant derived from dividing the equation by L which we can ignore. The amount and quality of the capital stock in relation to the labour force is an important determinant of labour productivity. Labour productivity will rise if the capital stock increases but it must also be remembered that because of diminishing marginal product the rate of increase in average

labour productivity will decline as the capital stock per worker increases. Economic growth, therefore, is dependent on the capital stock per worker and on average labour productivity.

How Does the Stock of Physical Capital Grow? Think back to Chapter 20 when we introduced a more complex circular flow diagram of the economy. We introduced withdrawals from and injections to the circular flow. When households receive income not all of it is spent, some of it is saved. The proportion of income saved is called the savings ratio. The Solow–Swan model noted the importance of the savings ratio in determining investment. Investment is spending on physical capital in the economy – on future productive capacity. Investment is financed by saving (we will look at the relationship between investment and saving in more detail in later chapters).

The level of investment in an economy is an important factor in determining the capital/labour ratio. Over time, economies will experience increases in population and in addition the existing capital stock will be subject to depreciation – machines wear out or become obsolete, plant will rust or break and so will need replacing. If capital is not replaced and the population increased, the capital/labour ratio will decline. Investment spending can be classified in two ways, gross investment and net investment. **Gross investment** is the total amount spent on capital stock in a period; **net investment** is spending on the capital stock minus the amount spent on replacing existing stock – in other words, spending on new capital and not on depreciation. The depreciation rate and the rate of growth of population will be key factors in determining the investment required to keep the capital/labour ratio constant. If the rate of depreciation of capital in an economy was 5 per cent a year and the population was rising at a rate of 1.5 per cent a year, then the economy would need investment to increase at a rate of 6.5 per cent a year to keep the capital/labour ratio constant.

gross investment the total spending on capital stock per period of time in an economy

net investment spending on the capital stock taking into account spending on depreciation of the existing capital stock

The ratio of capital to labour is one aspect, therefore, of how productive labour can be. One of the reasons put forward for the rapid growth in countries like Korea and Vietnam is that not only have they a supply of relatively cheap labour but both countries have invested in capital and as a result labour productivity is relatively high. Compare this to the various African countries we saw in Table 22.1 where average incomes remain very low, and the main difference is that in these African countries investment in capital stock is extremely low and so labour productivity is also very low.

We will now look further at how productivity of factors of production is determined.

How Productivity is Determined

Let us return to our example of Fernando. Although productivity is uniquely important in determining Fernando's standard of living, many factors determine Fernando's productivity. Fernando will be better at catching fish, for instance, if he has more fishing rods, if he has been trained in the best fishing techniques, if his island has a plentiful fish supply, and if he invents a better fishing lure. Each of these determinants of Fernando's productivity – which we can call *physical capital*, *human capital*, *natural resources* and *technological knowledge* – has a counterpart in more complex and realistic economies. Let's consider each of these factors in turn.

Physical Capital As we have seen, workers are more productive if the capital stock is high. The stock of equipment and structures that are used to produce goods and services is called **physical capital**, or just *capital*. For example, when carpenters make furniture, they use saws, lathes and drill presses. More tools allow work to be done more quickly and more accurately. That is, a carpenter with only basic hand tools can make less furniture each week than a carpenter with sophisticated and specialized woodworking equipment.

physical capital the stock of equipment and structures that are used to produce goods and services

An important feature of capital is that it is a *produced* factor of production. That is, capital is an input into the production process that in the past was an output from the production process. The carpenter uses a lathe to make the leg of a table. Earlier the lathe itself was the output of a firm that manufactures lathes. The lathe manufacturer in turn used other equipment to make its product. Thus, capital is a factor of production used to produce all kinds of goods and services, including more capital.

Human Capital A second determinant of productivity is human capital. Human capital is the economist's term for the knowledge and skills that workers acquire through education, training and experience. Human capital includes the skills accumulated in early childhood programmes, primary school, secondary school, university or college, and on-the-job training for adults in the labour force.

Although education, training and experience are less tangible than lathes, bulldozers and buildings, human capital is like physical capital in many ways. Like physical capital, human capital raises a nation's ability to produce goods and services. Also like physical capital, human capital is a produced factor of production. Producing human capital requires inputs in the form of teachers, lecturers, libraries and student time. Indeed, students can be viewed as 'workers' who have the important job of producing the human capital that will be used in future production.

Natural Resources A third determinant of productivity is **natural resources**. Natural resources are inputs into production that are provided by nature, such as land, rivers and mineral deposits. Natural resources take two forms: renewable and non-renewable. A forest is an example of a renewable resource. When one tree is cut down, a seedling can be planted in its place to be harvested in the future. Oil is an example of a non-renewable resource. Because oil is produced by nature over many thousands or even millions of years, there is only a limited supply. Once the supply of oil is depleted, it is impossible to create more (at least not for thousands of years).

natural resources the inputs into the production of goods and services that are provided by nature, such as land, rivers and mineral deposits

Differences in natural resources are responsible for some of the differences in standards of living around the world. The historical success of the United States was driven in part by the large supply of land well suited for agriculture. Today, some countries in the Middle East, such as Kuwait and Saudi Arabia, are rich simply because they happen to be on top of some of the largest pools of oil in the world.

Although natural resources can be important, they are not necessary for an economy to be highly productive in producing goods and services. Japan, for instance, is one of the richest countries in the world, despite having few natural resources. International trade makes Japan's success possible. Japan imports many of the natural resources it needs, such as oil, and exports its manufactured goods to economies rich in natural resources.

Technological Knowledge A fourth determinant of productivity is **technological knowledge** – the understanding of the best ways to produce goods and services. (Technology is defined as the application of knowledge to the environment to enable people to exercise greater control over that environment.) A hundred years ago, most Europeans and North Americans worked on farms, because farm technology required a high input of labour in order to feed the entire population. Today, thanks to advances in the technology of farming, a small fraction of the populations of western Europe, the USA and Canada can produce enough food to feed their entire population. This technological change made labour available to produce other goods and services.

technological knowledge society's understanding of the best ways to produce goods and services

Technological knowledge takes many forms. Some technology is common knowledge – after it starts to be used by one person, everyone becomes aware of it. For example, once Henry Ford successfully introduced production in assembly lines in the USA, other car makers and industrial producers throughout

the world quickly followed suit. Other technology is proprietary – it is known only by the company that discovers it. Only the Coca-Cola Company, for instance, knows the secret recipe for making its famous soft drink. Still other technology is proprietary for a short time. When a pharmaceutical company discovers a new drug, the patent system gives that company a temporary right to be its exclusive manufacturer. When the patent expires, however, other companies are allowed to make the drug. All these forms of technological knowledge are important for the economy's production of goods and services.

It is worthwhile to distinguish between technological knowledge and human capital. Although they are closely related, there is an important difference. Technological knowledge refers to *society's* understanding about how the world works. Human capital refers to the resources expended transmitting this understanding to the *labour force*. To use a relevant metaphor, knowledge is the quality of society's textbooks, whereas human capital is the amount of time that the population has devoted to reading them. Workers' productivity depends on both the quality of textbooks they have available and the amount of time they have spent studying them.

We have noted that investment is required in order to maintain the capital/labour ratio but in order to achieve growth, it is not simply sufficient to keep things in a constant state – this is where technical progress is so important. Technical progress means that the *quality* of physical and human capital is improved so for any given quantity of capital and labour, the average productivity of both is higher so that a higher output can be produced from the economy's factors of production. In essence, technical progress can help to counterbalance the effects of diminishing marginal product. If additional factors are employed but the total productivity of those factors increases then the economy can experience growth.

The production function can be written as:

$$\frac{Y}{L} = \beta f\left(\frac{K}{L}, 1\right)$$

where β (beta) is the rate of technological progress. Even if the capital/labour ratio is constant, this factor productivity can increase by whatever the increase in technological progress is. For example, if technological progress increases by 0.5 per cent a year, then $\beta = (1 + 0.005)$. Average labour productivity can increase by 0.5 per cent even if the ratio of capital to labour stays constant. If this is the case then the production function will actually shift upwards (or the production possibilities curve we studied in Chapter 19) will shift outwards reflecting the ability of the economy to produce more of every good.

We can summarize this analysis through the use of the following production function:

$$Y = \beta f(K, L, H, N)$$

Where Y denotes the quantity of output, K the quantity of physical capital, L the quantity of labour, H the quantity of human capital and N the quantity of natural resources. β is a variable that reflects the available production technology. As technology improves, β rises, so the economy produces more output from any given combination of inputs.

Many production functions have a property called *constant returns to scale*. If a production function has constant returns to scale, then a doubling of all the inputs causes the amount of output to double as well. Mathematically, we write that a production function has constant returns to scale if, for any positive number x :

$$xY = \beta f(xK, xL, xH, xN)$$

A doubling of all inputs is represented in this equation by $x = 2$. The right-hand side shows the inputs doubling, and the left-hand side shows output doubling.

If we set $x = 1/L$, the equation above becomes:

$$Y/L = \beta f\left(1, \frac{K}{L}, \frac{H}{L}, \frac{N}{L}\right)$$

Remember that Y/L is output per worker, a measure of productivity. This equation says that productivity depends on physical capital per worker ($\frac{K}{L}$), human capital per worker ($\frac{H}{L}$), and natural resources per worker ($\frac{N}{L}$). Productivity also depends on the state of technology, as reflected by the variable β . Thus, this equation provides a mathematical summary of the four determinants of productivity that we have just discussed.

SELF TEST List and describe four determinants of a country's productivity.

CASE STUDY Are Natural Resources a Limit to Growth?

Today the world's population is over 6 billion, about four times what it was a *century* ago. At the same time, many people are enjoying a much higher standard of living than did their great-grandparents. A perennial debate concerns whether this growth in population and living standards can continue in the future.

Many commentators have argued that natural resources provide a limit to how much the world's economies can grow. At first, this argument might seem hard to ignore. If the world has only a fixed supply of non-renewable natural resources, how can population, production and living standards continue to grow over time? Eventually, won't supplies of oil and minerals start to run out? When these shortages start to occur, won't they stop economic growth and, perhaps, even force living standards to fall?

Despite the apparent appeal of such arguments, most economists are less concerned about such limits to growth than one might imagine. They argue that technological progress often yields ways to avoid these limits. If we compare the economy today to the economy of the past, we see various ways in which the use of natural resources has improved. Modern cars have better petrol mileage. New houses have better insulation and require less energy to heat and cool them. More efficient oil rigs waste less oil in the process of extraction. Recycling allows some non-renewable resources to be reused. The development of alternative fuels, such as ethanol instead of petrol, allows us to substitute renewable for non-renewable resources.

Copper prices fluctuate over short periods of time but over longer periods of time are relatively stable. This might imply that supplies from mines like these are not running out as quickly as some people have predicted.



Fifty years ago, some conservationists were concerned about the excessive use of tin and copper. At the time, these were crucial commodities: tin was used to make many food containers, and copper was used to make telephone wire. Some people advocated mandatory recycling and rationing of tin and copper so that supplies would be available for future generations. Today, however, plastic has replaced tin as a material for making many food containers, and telephone calls often travel over fibre-optic cables, which are made from sand. Technological progress has made once-crucial natural resources less necessary.

But are all these efforts enough to permit continued economic growth? One way to answer this question is to look at the prices of natural resources. In a market economy, scarcity is reflected in market prices. If the world were running out of natural resources, then the prices of those resources would be rising over time. But, in fact, the opposite is more nearly true. Natural resource prices exhibit substantial short-run price fluctuations, but over long spans of time, the prices of most natural resources (adjusted for overall inflation) are stable or falling. It appears that our ability to conserve these resources is growing more rapidly than their supplies are dwindling. Market prices give no reason to believe that natural resources are a limit to economic growth.

ECONOMIC GROWTH AND PUBLIC POLICY

So far, we have determined that a society's standard of living depends on its ability to produce goods and services and that its productivity depends on physical capital, human capital, natural resources and technological knowledge. Let's now turn to the question faced by policymakers around the world: what can government policy do to raise productivity and living standards?

The Importance of Saving and Investment

We introduced the role of the savings ratio in relation to investment earlier in the chapter. Investment is necessary to sustain and increase the capital stock. Because capital is a produced factor of production, a society can change the amount of capital it has. If today the economy produces a large quantity of new capital goods, then tomorrow it will have a larger stock of capital and be able to produce more of all types of goods and services. Thus, one way to raise future productivity is to invest more current resources in the production of capital.

One of the *Ten Principles of Economics* is that people face trade-offs. This principle is especially important when considering the accumulation of capital. Because resources are scarce, devoting more resources to producing capital requires devoting fewer resources to producing goods and services for current consumption. That is, for society to invest more in capital, it must consume less and save more of its current income (the savings ratio must rise). The growth that arises from capital accumulation is not a free lunch: it requires that society sacrifice consumption of goods and services in the present in order to enjoy higher consumption in the future.

In the 1930s, the rulers of Russia deliberately diverted resources to the production of capital goods in an attempt to try to catch up with the richer and more industrialized Western countries such as Germany, the United States and the United Kingdom. It managed to expand production in core industries such as coal and steel significantly which in turn increased its capacity to produce other capital goods (and military equipment). The trade-off for the Russian people at the time was fewer consumer goods and a harsh life. It could be argued that whilst this represented significant short-term hardship for many of its people, the decision helped Russia to be in a position to fight off the Germans in World War II and emerge to be one of the planet's superpowers.

In later chapters we examine in more detail how the economy's financial markets coordinate saving and investment. We also examine how government policies influence the amount of saving and investment that takes place. At this point it is important to note that encouraging saving and investment is one way that a government can encourage growth and, in the long run, raise the economy's standard of living.

To see the importance of investment for economic growth, consider Figure 22.2, which displays data on 19 countries. Panel (a) shows each country's average growth rate between 1961 and 2011. The countries are ordered by their average annual growth rates, from most to least rapid. Panel (b) shows the percentage of GDP that each country devoted to investment as an average over the period 1961–2011. Note that some data for some countries in both cases are not complete – there are gaps in data for Afghanistan, in particular.

The correlation between growth and investment is strong. Countries that devote a large share of GDP to investment, such as China, Japan and Australia, also have a stronger average growth rate. Countries that devote a small share of GDP to investment, such as the Central African Republic, Zimbabwe and Bangladesh, tend to have low growth rates. Drilling down more deeply into the figures shows that there are a number of years where investment in the Central African Republic and Bangladesh has been in single digits, whereas in China, investment began to increase from around 1978 to over 30 per cent and in the early 1990s to over 44 per cent. Investment hovered in the upper 30 per cents until the early part of the new century when investment was consistently above 44 per cent and reached over 48 per cent from 2009. In the 1970s and 1980s, Japan consistently allocated over 30 per cent of GDP to investment and growth during those years was generally high at around 5.0 per cent. In the last ten years, Japan has experienced considerable economic problems, its growth rates have stagnated but so has the proportion of GDP accounted for by investment – under 20 per cent from 2009 onwards.

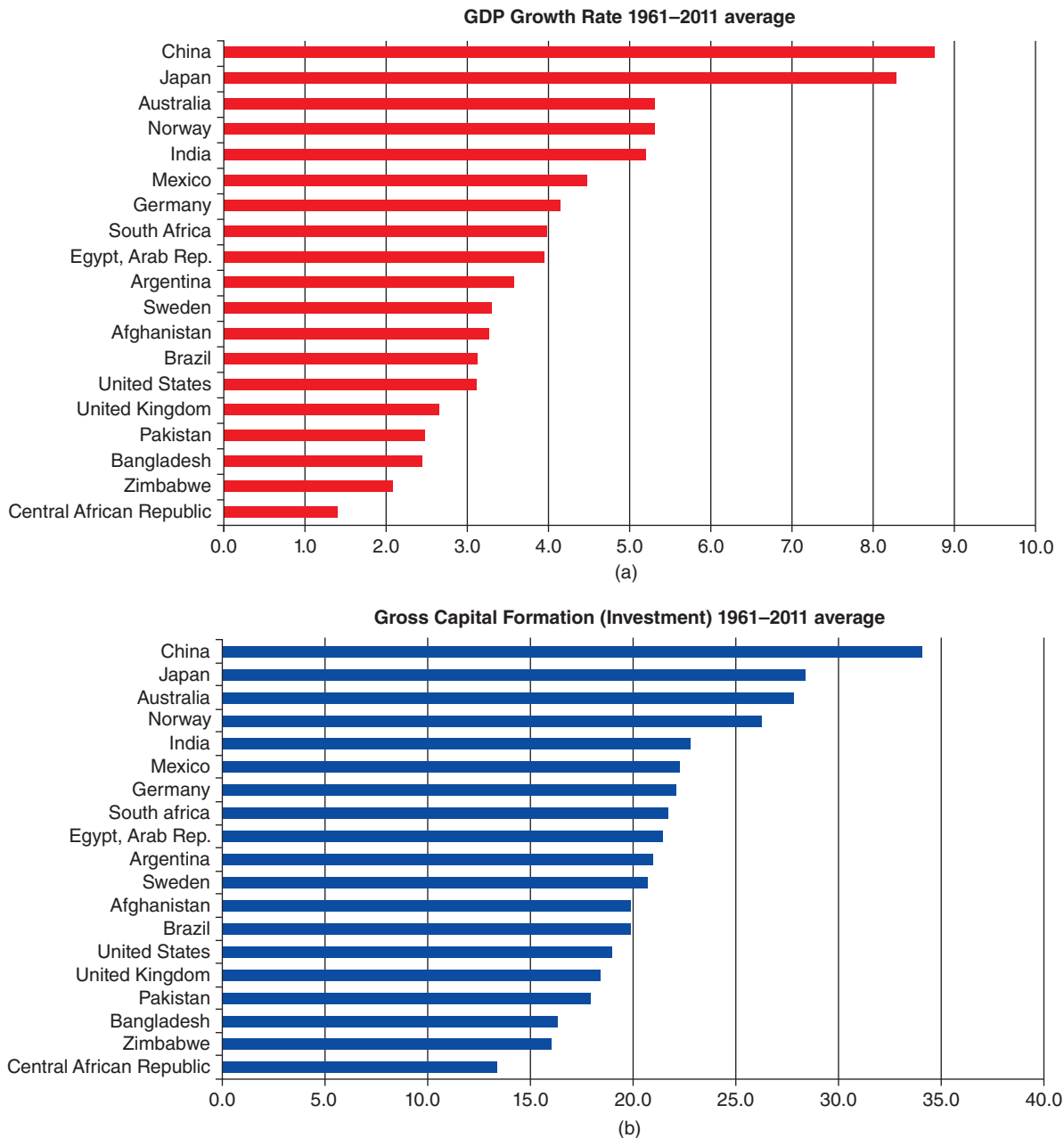
Studies that examine a more comprehensive list of countries confirm this strong correlation between investment and growth. It should also be noted that the figures can be affected by the starting base; countries such as China now devote over a third of GDP to investment whereas in the early 1960s it hovered between 10.5 per cent and 22.08 per cent.

There is, however, a problem in interpreting these data. As the appendix to Chapter 2 discussed, a correlation between two variables does not establish which variable is the cause and which is the effect. It is possible that high investment causes high growth, but it is also possible that high growth causes high investment. (Or, perhaps, high growth and high investment are both caused by a third variable that

has been omitted from the analysis.) The data by themselves cannot tell us the direction of causation. Nevertheless, because capital accumulation affects productivity so clearly and directly, many economists interpret these data as showing that high investment leads to more rapid economic growth.

FIGURE 22.2**Growth and Investment**

Panel (a) shows the average growth rate of GDP for 19 countries over the period 1961–2011. Panel (b) shows the average percentage of GDP that each country devoted to investment over this period. The figure suggests that investment and growth are positively correlated.



Diminishing Returns and the Catch-Up Effect

Suppose that a government, convinced by the evidence in Figure 22.2, pursues policies that raise the nation's saving rate. What happens? With the nation saving more, fewer resources are needed to make consumption goods, and more resources are available to make capital goods. As a result, the capital stock increases, leading to rising productivity and more rapid growth in GDP. But how long does this higher rate of growth last? Assuming that the saving rate remains at its new higher level, does the growth rate of GDP stay high indefinitely or only for a period of time?

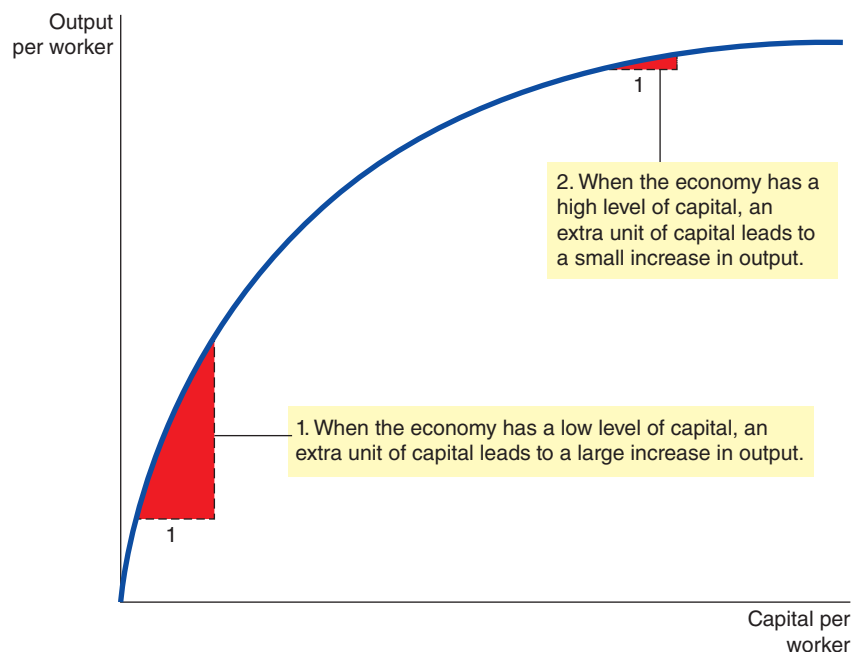
We know that the traditional view of the production process is that capital is subject to diminishing returns: as the stock of capital rises, the extra output produced from an additional unit of capital falls. In other words, when workers already have a large quantity of capital to use in producing goods and services, giving them an additional unit of capital increases their productivity only slightly. This is illustrated in Figure 22.3, which shows how the amount of capital per worker determines the amount of output per worker, holding constant all the other determinants of output. Because of diminishing returns, an increase in the saving rate leads to higher growth only for a while. As the higher saving rate allows more capital to be accumulated, the benefits from additional capital become smaller over time, and so growth slows down. In the long run, the higher saving rate leads to a higher level of productivity and income, but not to higher growth in these variables. Reaching this long run, however, can take quite a while. According to studies of international data on economic growth, increasing the saving rate can lead to substantially higher growth for a period of several decades.

The diminishing returns to capital has another important implication: other things equal, it is easier for a country to grow fast if it starts out relatively poor.

FIGURE 22.3

Illustrating the Production Function

This figure shows how the amount of capital per worker influences the amount of output per worker. Other determinants of output, including human capital, natural resources and technology are held constant. The curve becomes flatter as the amount of capital increases because of diminishing returns to capital.



This effect of initial conditions on subsequent growth is sometimes called the **catch-up effect**. In poor countries, workers lack even the most rudimentary tools and, as a result, have low productivity. Small amounts of capital investment would substantially raise these workers' productivity. By contrast, workers in rich countries have large amounts of capital with which to work, and this partly explains their high productivity. Yet with the amount of capital per worker already so high, additional capital investment has a relatively small effect on productivity. Studies of international data on economic growth confirm this catch-up effect: controlling for other variables, such as the percentage of GDP devoted to investment, poor countries do tend to grow at a faster rate than rich countries.

catch-up effect the property whereby countries that start off poor tend to grow more rapidly than countries that start off rich

This catch-up effect can help explain variations in observations on average investment as a proportion of GDP and average growth rates over time. Between 1961 and 2011, Japan allocated around 10 per cent more to investment as a proportion of GDP compared to China, but China grew at an average rate 1.7 times the Japanese annual average growth rate. The explanation is the catch-up effect. In 1961 Japan had GDP per person about 15 times that of China, in part because investment in China up to the end of the 1960s had been so low. With a small initial capital stock, the benefits to capital accumulation were much greater in China, and this gave China a higher subsequent growth rate.

Investment from Abroad

So far we have discussed how policies aimed at increasing a country's saving rate can increase investment and, thereby, long-term economic growth. Yet saving by domestic residents is not the only way for a country to invest in new capital. The other way is investment by foreigners.

Investment from abroad takes several forms. BMW might build a car factory in Portugal. A capital investment that is owned and operated by a foreign entity is called **foreign direct investment**. Alternatively, a German might buy equity in a Portuguese corporation (that is, buy a share in the ownership of the corporation); the Portuguese corporation can use the proceeds from the equity sale to build a new factory. An investment that is financed with foreign money but operated by domestic residents is called **foreign portfolio investment**. In both cases, Germans provide the resources necessary to increase the stock of capital in Portugal. That is, German saving is being used to finance Portuguese investment.

foreign direct investment capital investment that is owned and operated by a foreign entity

foreign portfolio investment investment that is financed with foreign money but operated by domestic residents

When foreigners invest in a country, they do so because they expect to earn a return on their investment. BMW's car factory increases the Portuguese capital stock and, therefore, increases Portuguese productivity and Portuguese GDP. Yet BMW takes some of this additional income back to Germany in the form of profit. Similarly, when a German investor buys Portuguese equity, the investor has a right to a portion of the profit that the Portuguese corporation earns.

Investment from abroad, therefore, does not have the same effect on all measures of economic prosperity. Recall that GDP is the income earned within a country by both residents and non-residents, whereas GNP is the income earned by residents of a country both at home and abroad. When BMW opens its car factory in Portugal, some of the income the factory generates accrues to people who do not live in Portugal. As a result, foreign investment in Portugal raises the income of the Portuguese (measured by GNP) by less than it raises the production in Portugal (measured by GDP).

Nevertheless, investment from abroad is one way for a country to grow. Even though some of the benefits from this investment flow back to the foreign owners, this investment does increase the economy's stock of capital, leading to higher productivity and higher wages. Moreover, investment from abroad is one way for poorer countries to learn the state-of-the-art technologies developed and used in richer countries. For these reasons, many economists who advise governments in less developed economies advocate

policies that encourage investment from abroad. Often this means removing restrictions that governments have imposed on foreign ownership of domestic capital.

An organization that tries to encourage the flow of capital to poor countries is the World Bank. This international organization obtains funds from the world's advanced industrialized countries, and uses these resources to make loans to less developed countries so that they can invest in roads, sewage systems, schools and other types of capital. It also offers the countries advice about how the funds might best be used. The World Bank, together with its sister organization, the International Monetary Fund (IMF), was set up after World War II. One lesson from the war was that economic distress often leads to political turmoil, international tensions and military conflict. Thus, every country has an interest in promoting economic prosperity around the world. The World Bank and the IMF were set up to achieve that common goal.

Education

Education – investment in human capital – is at least as important as investment in physical capital for a country's long-run economic success. In the developed economies of Western Europe and North America, each extra year of schooling raises a worker's income by about 10 per cent on average. In less developed countries, where human capital is especially scarce, the gap between the wages of educated and uneducated workers is even larger. Thus, one way in which government policy can enhance the standard of living is to provide good schools and to encourage the population to take advantage of them.

Investment in human capital, like investment in physical capital, has an opportunity cost. When students are in school, they forgo the wages they could have earned. In less developed countries, children often drop out of school at an early age, even though the benefit of additional schooling is very high, simply because their labour is needed to help support the family.

Some economists have argued that human capital is particularly important for economic growth because human capital conveys positive externalities. An *externality*, remember, is the effect of one person's actions on the well-being of a third party. An educated person, for instance, might generate new ideas about how best to produce goods and services. If these ideas enter society's pool of knowledge, so that everyone can use them, then the ideas are an external benefit of education. In this case, the return to schooling for society is even greater than the return for the individual. This argument would justify the large subsidies to human capital investment that we observe in the form of public education.

One problem facing some poor countries is the **brain drain** – the emigration of many of the most highly educated workers to rich countries where these workers can enjoy a higher standard of living. If human capital does have positive externalities, then this brain drain makes those people left behind poorer than they otherwise would be. This problem offers policymakers a dilemma. On the one hand, rich countries like those of western Europe and North America have the best systems of higher education, and it would seem natural for poor countries to send their best students abroad to earn higher degrees. On the other hand, those students who have spent time abroad may choose not to return home, and this brain drain will reduce the poor nation's stock of human capital even further.

brain drain the emigration of many of the most highly educated workers to rich countries

Health and Nutrition

The term *human capital* usually refers to education, but it can also be used to describe another type of investment in people: expenditures that lead to a healthier population. Other things equal, healthier workers are more productive. The right investments in the health of the population provide one way for a nation to increase productivity and raise living standards.

Economic historian, Robert Fogel, has suggested that a significant factor in long-run economic growth is improved health from better nutrition. He estimates that in Great Britain in 1780 about one in five people were so malnourished that they were incapable of manual labour. Among those who could work, insufficient intake of calories substantially reduced the work effort they could carry out. As nutrition improved so did workers' productivity.

Fogel studies these historical trends in part by looking at the height of the population. Short stature can be an indicator of malnutrition, especially during gestation and the early years of life. Fogel finds that as nations develop economically, people eat more and the population gets taller. From 1775 to 1975, the average caloric intake in Great Britain rose by 26 per cent and the height of the average man rose by 3.6 inches (around 10 cm). Similarly, during the spectacular economic growth in the Republic of Korea from 1962 to 1995, caloric consumption rose by 44 per cent and, average male height rose by 2 inches (5 cm). Of course, a person's height is determined by a combination of genetic predisposition and environment. But because the genetic make-up of a population is slow to change, such increases in average height are likely due to changes in the environment – nutrition being the obvious explanation.

Moreover, studies have found that height is an indicator of productivity. Looking at data on a large number of workers at a point in time, researchers have found that taller workers tend to earn more. Because wages reflect a worker's productivity, this finding suggests that taller workers tend to be more productive. The effect of height on wages is especially pronounced in poorer countries, where malnutrition is a bigger risk.

Fogel won the Nobel Prize in Economics in 1993 for his work in economic history, which includes not only his studies on nutrition but also work on American slavery and the role of the railroads in the development of the American economy. In the lecture he gave when awarded the Prize, he surveyed the evidence on health and economic growth. He concluded that 'improved gross nutrition accounts for roughly 30 per cent of the growth of per capita income in Britain between 1790 and 1980'.

Today malnutrition is fortunately rare in developed nations (obesity is a more widespread problem). But for people in developing nations, poor health and inadequate nutrition remain obstacles to higher productivity and improved living standards. The United Nations estimates that almost a third of the population in sub-Saharan Africa is undernourished.

The causal link between health and wealth runs in both directions. Poor countries are poor, in part, because their populations are not healthy, and their populations are not healthy, in part, because they are poor and cannot afford adequate health care and nutrition. It is a vicious circle. But this fact opens the possibility of a virtuous circle. Policies that lead to more rapid economic growth would naturally improve health outcomes, which in turn would further promote economic growth.

Property Rights, Political Stability and Good Governance

Other ways in which policymakers can foster economic growth are by protecting property rights, promoting political stability and maintaining good governance. As noted when we discussed economic interdependence, production in market economies arises from the interactions of millions of individuals and firms. When you buy a car, for instance, you are buying the output of a car dealer, a car manufacturer, a steel company, an iron ore mining company and so on. This division of production among many firms allows the economy's factors of production to be used as effectively as possible. To achieve this outcome, the economy has to coordinate transactions among these firms, as well as between firms and consumers. Market economies achieve this coordination through market prices. That is, market prices are the instrument with which the invisible hand of the marketplace brings supply and demand into balance.

An important prerequisite for the price system to work is an economy-wide respect for *property rights*. A mining company will not make the effort to mine iron ore if it expects the ore to be stolen. The company mines the ore only if it is confident that it will benefit from the ore's subsequent sale. For this reason, courts serve an important role in a market economy: they enforce property rights. Through the criminal justice system, the courts discourage direct theft. In addition, through the civil justice system, the courts ensure that buyers and sellers live up to their contracts.

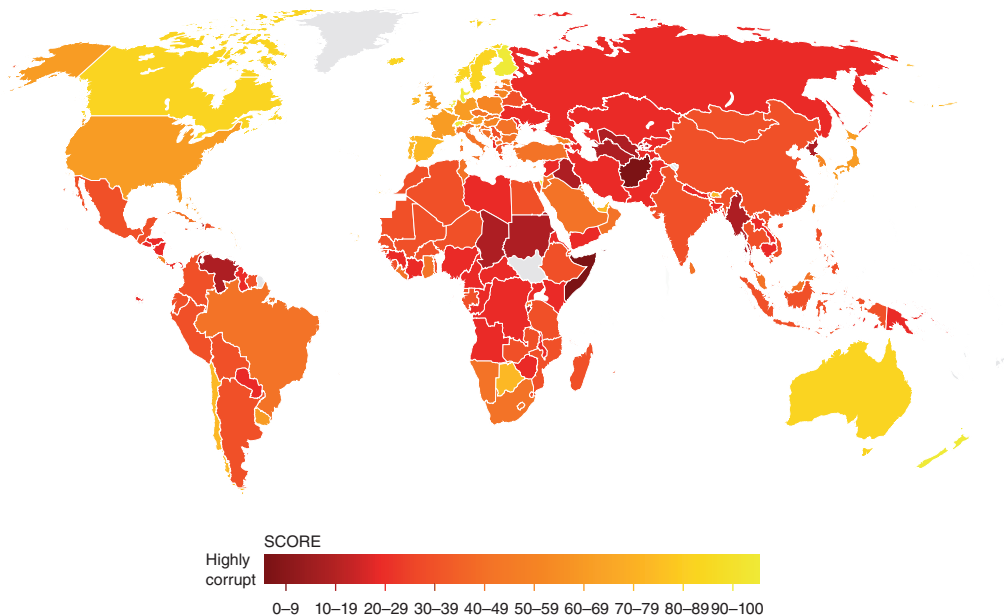
Although people in developed countries tend to take property rights for granted, those living in less developed or emerging countries understand that lack of property rights can be a major problem. In many countries, the system of justice does not work well. Contracts are hard to enforce, and fraud often goes unpunished. In more extreme cases, the government not only fails to enforce property rights but actually infringes upon them. To do business in some countries, firms are expected to bribe powerful government officials. Such corruption impedes the coordinating power of markets. It also discourages domestic saving

and investment from abroad. The problem of corruption and bribery is extensive. Transparency International is a pressure group that exists to advance the cause of anti-corruption. Each year it publishes an index showing perceptions of corruption across the world. Countries are scored on a scale of 0–100 with 0 being highly corrupt and 10 being highly clean. Around two-thirds of the 176 countries in the ranking fall below 50. The extent of the problem is highlighted in the map produced by Transparency International reproduced in Figure 22.4.

FIGURE 22.4

Corruption Perceptions Index 2012

The map shows the perceptions of corruption across 176 countries around the world. The darker the red on the map, the more there is a perception of corruption and the more yellow, the cleaner the perception.



One threat to property rights is political instability. When revolutions and coups are common, there is doubt about whether property rights will be respected in the future. If a revolutionary government might confiscate the capital of some businesses, as was often true after communist revolutions, domestic residents have less incentive to save, invest and start new businesses. At the same time, foreigners have less incentive to invest in the country. Even the threat of revolution can act to depress a nation's standard of living. It is no coincidence that countries with a strong military power, who are subject to frequent coups, are ones that are at the bottom of any standard of living league table.

Thus, economic prosperity depends in part on political prosperity. A country with an efficient court system, honest government officials and a stable constitution will enjoy a higher economic standard of living than a country with a poor court system, corrupt officials and frequent revolutions and coups. These are the key features of good governance – the extent to which a country is ruled by sound democracies, where the rule of law, the authority of law, the absence of corruption and independent judicial processes are in existence. If these things are in place then it means that contracts and property rights can be enforced and free markets can operate effectively to allocate scarce resources. Without good governance, many economists believe that economic development will be compromised.

Free Trade

Some of the world's poorest countries have tried to achieve more rapid economic growth by pursuing *inward-oriented policies*. These policies are aimed at raising productivity and living standards within the country by avoiding interaction with the rest of the world. This approach gets support from some domestic firms, which claim that they need protection from foreign competition in order to compete and grow. This infant-industry argument, together with a general distrust of foreigners, has at times led policymakers in less developed countries to impose tariffs and other trade restrictions.

Most economists today believe that poor countries are better off pursuing *outward-oriented policies* that integrate these countries into the world economy. When we studied international trade earlier in the book, we showed how international trade can improve the economic well-being of a country's citizens. Trade is, in some ways, a type of technology. When a country exports wheat and imports steel, the country benefits in the same way as if it had invented a technology for turning wheat into steel. A country that eliminates trade restrictions will, therefore, experience the same kind of economic growth that would occur after a major technological advance.

The adverse impact of inward orientation becomes clear when one considers the small size of many less developed economies. The total GDP of Argentina, for instance, is about that of the southeast of England (if we include London). Imagine what would happen if the southeast were suddenly to declare that it was illegal to trade with anyone beyond the regional boundaries. Without being able to take advantage of the gains from trade, the southeast would need to produce all the goods it consumes. It would also have to produce all its own capital goods, rather than importing state-of-the-art equipment from other cities. Living standards in the southeast would fall immediately, and the problem would probably only get worse over time. This is precisely what happened when Argentina pursued inward-oriented policies throughout much of the 20th century and can also partly explain the economic problems in North Korea which has effectively cut itself off from the rest of the world. By contrast, countries pursuing outward-oriented policies, such as the Republic of Korea, Singapore and Taiwan, have enjoyed high rates of economic growth.

The amount that a nation trades with others is determined not only by government policy but also by geography. Countries with good natural seaports find trade easier than countries without this resource. It is not a coincidence that many of the world's major cities, such as Paris, New York, Hong Kong and London, are located either next to oceans or on the banks of a major river giving easy access for seafaring trade vessels. Similarly, because landlocked countries find international trade more difficult, they tend to have lower levels of income than countries with easy access to the world's waterways.

Research and Development

The primary reason that living standards are higher today than they were a century ago is that technological knowledge has advanced. The telephone, the transistor, the computer and the internal combustion engine, are among the thousands of innovations that have improved the ability to produce goods and services.

Although most technological advance comes from private research by firms and individual inventors, there is also a public interest in promoting these efforts. To a large extent, knowledge is a *public good*: once one person discovers an idea, the idea enters society's pool of knowledge and other people can freely use it (subject to any legal restrictions such as those imposed by intellectual property rights). Just as government has a role in providing a public good such as national defence, it also has a role in encouraging the research and development of new technologies. The governments in most advanced countries do this in a number of ways, for example through science research laboratories owned and funded by the government, or through a system of research grants offered to promising researchers. It may also offer tax breaks and concessions for firms engaging in research and development.

Yet another way in which government policy encourages research is through the patent system. When a person or firm invents a new product, such as a new drug, the inventor can apply for a patent. If the product is deemed truly original, the government awards the patent, which gives the inventor the exclusive right to make the product for a specified number of years. In essence, the patent gives the inventor a property right over his invention, turning his new idea from a public good into a private good. By allowing inventors to profit from their inventions – even if only temporarily – the patent system enhances the incentive for individuals and firms to engage in research.

Population Growth

Economists and other social scientists have long debated how population growth affects a society. The most direct effect is on the size of the labour force: a large population means more workers to produce goods and services. At the same time, it means more people to consume those goods and services. Beyond these obvious effects, population growth interacts with the other factors of production in ways that are less obvious and more open to debate.

Stretching Natural Resources Thomas Robert Malthus (1766–1834), an English minister and early economic thinker, is famous for his book entitled *An Essay on the Principle of Population as It Affects the Future Improvement of Society*. In it, Malthus offered what may be history's most chilling forecast. Malthus argued that an ever-increasing population would continually strain society's ability to provide for itself. As a result, mankind was doomed to forever live in poverty.

Malthus's logic was simple. He began by noting that 'food is necessary to the existence of man' and that 'the passion between the sexes is necessary and will remain nearly in its present state'. He concluded that 'the power of population is infinitely greater than the power in the earth to produce subsistence for man'. According to Malthus, the only check on population growth was 'misery and vice'. Attempts by charities or governments to alleviate poverty were counterproductive, he argued, because they merely allowed the poor to have more children, placing even greater strains on society's productive capabilities.

Fortunately, Malthus's dire forecast was far off the mark. Although the world population has increased about six-fold over the past two centuries, living standards around the world are on average much higher. As a result of economic growth, chronic hunger and malnutrition are less common now than they were in Malthus's day. Famines occur from time to time, but they are more often the result of an unequal income distribution or political instability than an inadequate production of food.

Where did Malthus go wrong? He assumed – correctly – that world population would rise exponentially, since as more people are born and survive, more and more people are born as their children, and so on. But he also assumed – incorrectly – that the amount of food produced could rise only linearly, by increasing the amount of land under cultivation with productivity remaining constant. But growth in mankind's ingenuity over the years has offset the effects of a larger population. Pesticides, fertilizers, mechanized farm equipment, new crop varieties and other technological advances that Malthus never imagined have allowed each farmer to feed ever greater numbers of people. Even with more mouths to feed, fewer farmers are necessary because each farmer is so much more productive than Malthus ever imagined.

Diluting the Capital Stock Whereas Malthus worried about the effects of population on the use of natural resources, some modern theories of economic growth emphasize its effects on capital accumulation. According to these theories, high population growth reduces GDP per worker because rapid growth in the number of workers forces the capital stock to be spread more thinly. In other words, when population growth is rapid, each worker is equipped with less capital. A smaller quantity of capital per worker leads to lower productivity and lower GDP per worker.

This problem is most apparent in the case of human capital. Countries with high population growth have large numbers of school-age children. This places a larger burden on the educational system. It is not surprising, therefore, that educational attainment tends to be low in countries with high population growth.

Thomas Robert Malthus



The differences in population growth around the world are large. In developed countries, such as the United States, the United Kingdom and the other countries of western Europe, the population has risen only about 1 per cent per year in recent decades and is expected to rise even more slowly in the future. In contrast, in many poor African countries, population grows at about 3 per cent per year. At this rate, the population doubles every 23 years (remember the rule of 70). This rapid population growth makes it harder to provide workers with the tools and skills they need to achieve high levels of productivity.

Although rapid population growth is not the main reason that less developed countries are poor, some analysts believe that reducing the rate of population growth would help these countries raise their standards of living. In some countries, this goal is accomplished directly with laws that regulate the number of children families may have. China, for instance, allows only one child per family living in urban areas; couples who violate this rule are subject to substantial fines. (In 2008, the National Population and Family Planning Commission announced that there were no planned changes to the policy for 'at least another decade'.) In countries with greater freedom, the goal of reduced population growth is accomplished less directly by increasing awareness of birth control techniques.

Another way in which a country can influence population growth is to apply one of the *Ten Principles of Economics*: people respond to incentives. Bearing a child, like any decision, has an opportunity cost. When the opportunity cost rises, people will choose to have smaller families. In particular, women with the opportunity to receive good education and desirable employment tend to want fewer children than those with fewer opportunities outside the home. Hence, policies that foster equal treatment of women are one way for less developed economies to reduce the rate of population growth and, perhaps, raise their standards of living.

Promoting Technological Progress Although rapid population growth may depress economic prosperity by reducing the amount of capital each worker has, it may also have some benefits. Some economists have suggested that world population growth has been an engine of technological progress and economic prosperity. The mechanism is simple: if there are more people, then the greater the probability that some of those people will come up with new ideas that will lead to technological progress, which benefits everyone.

Economist Michael Kremer has provided some support for this hypothesis in an article titled 'Population Growth and Technological Change: One Million BC to 1990', which was published in the *Quarterly Journal of Economics* in 1993. Kremer begins by noting that over the broad span of human history, world growth rates have increased as world population has. For example, world growth was more rapid when the world population was 1 billion (which occurred around the year 1800) than it was when the population was only 100 million (around 500 BC). This fact is consistent with the hypothesis that having more people induces more technological progress.

Kremer's second piece of evidence comes from comparing regions of the world. The melting of the polar ice caps at the end of the ice age around 10,000 BC flooded the land bridges and separated the world into several distinct regions that could not communicate with one another for thousands of years. If technological progress is more rapid when there are more people to discover things, then larger regions should have experienced more rapid growth.

According to Kremer, that is exactly what happened. The most successful region of the world in 1500 (when Columbus re-established technological contact) comprised the 'Old World' civilizations of the large Eurasia-Africa region. Next in technological development were the Aztec and Mayan civilizations in the Americas, followed by the hunter-gatherers of Australia, and then the primitive people of Tasmania, who lacked even fire-making and most stone and bone tools.

The smallest isolated region was Flinders Island, a tiny island between Tasmania and Australia. With the smallest population, Flinders Island had the fewest opportunities for technological advance and, indeed, seemed to regress. Around 3000 BC, human society on Flinders Island died out completely. A large population, Kremer concludes, is a prerequisite for technological advance.

At first sight this conclusion does seem to be at odds with casual empirical observation of the modern world: as we previously noted, in many rich, developed countries population growth has been only about 1 per cent per year in recent decades, while in many poor countries, such as those of sub-Saharan Africa, population growth is much higher. So why doesn't this higher population growth help these poor countries to grow, if Kremer's argument is right? The point is that Kremer was really analysing *world* economic

growth, or rather, economic growth in isolated regions of the world. Nowadays, in a very poor country, it is unlikely that technological advances will be made that are not already known in developed countries; the problem is not lack of technological progress but difficulty in applying technology because of the scarcity of human capital and perhaps because of problems arising from political instability and corruption, as we discussed earlier. Moreover, because many talented people from less developed countries tend to emigrate to richer, developed countries where they may work, for example, as scientists or entrepreneurs – a phenomenon we referred to earlier as the *brain drain* – population growth in less developed countries may actually enhance economic growth in developed countries.

SELF TEST Describe three ways in which a government policymaker can try to raise the growth in living standards in a society. Are there any drawbacks to these policies?

CONCLUSION: THE IMPORTANCE OF LONG-RUN GROWTH

In this chapter we have discussed what determines the standard of living in a nation and how policymakers can endeavour to raise the standard of living through policies that promote economic growth. Most of this chapter is summarized in one of the *Ten Principles of Economics*: a country's standard of living depends on its ability to produce goods and services. Policymakers who want to encourage growth in standards of living must aim to increase their nation's productive ability by encouraging rapid accumulation of the factors of production and ensuring that these factors are employed as effectively as possible.

Economists differ in their views of the role of government in promoting economic growth. At the very least, government can lend support to the invisible hand by maintaining property rights and political stability. More controversial is whether government should target and subsidize specific industries that might be especially important for technological progress. There is no doubt that these issues are among the most important in economics. The success of one generation's policymakers in learning and heeding the fundamental lessons about economic growth determines what kind of world the next generation will inherit.

IN THE NEWS

Economic Growth in China

Just as there were eyebrows raised for the way in which Japan transformed itself in the post-war period, in the last 20 years growth in China has been nothing short of remarkable. In Japan, economic growth has ground to a halt and the country is struggling with a number of problems. How long can China's growth continue?

How Sustainable is Economic Growth in China?

To many people outside China, the economic growth that has projected the country onto the world economic stage, as well as its political one, has been remarkable and

generated primarily by having a very large supply of labour willing to work for low wages to manufacture goods at low cost which have been bought by countries throughout the world. Professor Justin Yifu Lin, former Chief Economist and Senior Vice President of the World Bank,

has spoken at length about China's capacity to continue growing at high rates in a series of speeches at prominent higher education institutions including the London School of Economics in December 2012 where he spoke on 'Demystifying the Chinese Economy'.

(Continued)

Lin noted that in 1979 China had a per capita income of just \$182 which was less than a third of the average of Sub-Saharan African countries. Over the last 30 years economic growth has averaged over 9 per cent, 600 million people have been raised out of poverty and per capita income is now over \$4,000. In 2009, China became the world's second largest economy and many predict that it will become the largest in the next ten years. Lin referred to studies by economic historian, Angus Maddison (2001) which showed that average annual per capita income growth in the West was around 0.05 per cent prior to the 18th century, meaning it took 1,400 years for European per capita income to double, rising to 1 per cent in the 19th century (taking 70 years for per capita income to double) and 2 per cent in the 20th century (35 years for per capita income to double).

The industrial revolution led to millions of people moving from rural

Professor Justin Yifu Lin, former Chief Economist and Senior Vice President of the World Bank believes that the Chinese economy can continue to grow strongly but that it will have to change its focus as its comparative advantage changes.



occupations to manufacturing and infrastructure but the risks involved in the process were huge. China, Lin argues, has been able to take advantage of what he calls 'backwardness' – the ability to 'borrow' technology, innovation and institutions at relatively low risk and cost from advanced nations. The World Bank Commission noted this as the ability for developing countries to 'import what the rest of the world knew and export what it wanted'. This ability is one factor that can help a developing nation to secure sustained per-capita growth at high levels for a considerable period of time.

The early stages of China's growth were characterized by the setting up of heavy industry but in the context of very limited trade with the outside world. In the 1980s and beyond, the Chinese government began to exploit the advantages it had in labour supply, developing more modern industries and adopting some market-based reforms within the framework of a single-party state. The development of a more market-led economy enabled China to be able to become more outward-facing and export growth began to fuel economic growth.

Lin notes, however, that economic growth in China was not without its costs as rising inequality being a consequence. However, large companies and a rump of very rich people are able to access funds for investment at relatively favourable rates and because rich people and corporations tend to have a relatively high savings ratio, funds for further investment in improving the capital stock and productivity in China are available more freely than in more developed nations.

Some of these distortions will, however, have to be tackled by the Chinese government as it moves

ever further towards a fully functioning market economy. This includes the necessity to improve the quality of the banking sector to ensure finance is available for the development of small firms and for supplying credit to the consumer markets, tackle corporate governance and property rights and further improving infrastructure.

Lin believes that China can continue to grow at rates in excess of 8 per cent or more for the next 20 years. He bases this conclusion on the continued benefits the country can gain from 'backwardness' and the fact that there is still a massive gap to make up before China matches the per-capita income of country's like the United States. However, China will have to move to exploiting technological change rather than relying on its supply of cheap labour because this comparative advantage is already beginning to narrow as other low wage economies develop. China will need to move from absorbing technology, therefore, to being an innovator.

Questions

- 1 **China's population is in excess of 1 billion. Is this natural factor endowment a key reason why China has been able to generate such remarkable economic growth in the last 30 years? If so, why can't other countries with very large populations also grow at such rates?**
- 2 **Lin notes that 60 million people have been lifted out of poverty in China. The industrial revolution also meant millions across Europe were able to leave the grind of agricultural poverty and find work in new industries. Does such a transition come with its own challenges to a country like China?**

- 3** If China has been growing at an average of 9 per cent in the last ten years, how long will it take for per capita incomes to double?
- 4** Explain how the idea of 'backwardness' allows China to be able to generate rapid and sustained economic growth over a period.
- 5** Lin has warned that China cannot rely on its comparative advantage in low-wage labour and on backwardness for much longer. Why do you think he believes this and why has he urged the Chinese government to push China to pursue innovation?

SUMMARY

- Economic prosperity, as measured by GDP per person, varies substantially around the world. The average income in the world's richest countries is more than ten times that in the world's poorest countries. Because growth rates of real GDP also vary substantially, the relative positions of countries can change dramatically over time.
- The standard of living in an economy depends on the economy's ability to produce goods and services. Productivity, in turn, depends on the amounts of physical capital, human capital, natural resources and technological knowledge available to workers.
- Government policies can try to influence the economy's growth rate in many ways: by encouraging saving and investment, encouraging investment from abroad, fostering education, maintaining property rights and political stability, allowing free trade, promoting the research and development of new technologies, and controlling population growth.
- The accumulation of capital is subject to diminishing returns: the more capital an economy has, the less additional output the economy gets from an extra unit of capital. Because of diminishing returns, higher saving leads to higher growth for a period of time, but growth eventually slows down as the economy approaches a higher level of capital, productivity and income. Also because of diminishing returns, the return to capital is especially high in poor countries. Other things equal, these countries can grow faster because of the catch-up effect.

QUESTIONS FOR REVIEW

- 1 What does the level of a nation's GDP measure? What does the growth rate of GDP measure? Would you rather live in a nation with a high level of GDP and a low growth rate, or in a nation with a low level of GDP and a high growth rate?
- 2 What is the difference between the terms 'productivity' and 'production'?
- 3 List and describe four determinants of productivity.
- 4 In what way is a university degree a form of capital?
- 5 Explain how higher saving leads to a higher standard of living. What might deter a policymaker from trying to raise the rate of saving?
- 6 Does a higher rate of saving lead to higher growth temporarily or indefinitely?
- 7 Why is technological progress important in improving growth rates in any economy?
- 8 Why would removing a trade restriction, such as a tariff, lead to more rapid economic growth?
- 9 How does the rate of population growth influence the level of GDP per person?
- 10 Describe two ways in which the government can try to encourage advances in technological knowledge.

PROBLEMS AND APPLICATIONS

- 1 Most countries import substantial amounts of goods and services from other countries. Yet the chapter says that a nation can enjoy a high standard of living only if it can produce a large quantity of goods and services itself. Can you reconcile these two facts?

- 2 List the capital inputs necessary to produce each of the following:
 - a. cars
 - b. secondary education
 - c. air travel
 - d. fruit and vegetables.
- 3 UK income per person today is roughly four times what it was a century ago. Many other countries have also experienced significant growth over that period. What are some specific ways in which your standard of living is likely to differ from that of your great-grandparents?
- 4 The chapter discusses how employment in developed economies has declined relative to output in the agricultural sector. Can you think of another sector of the economy where the same phenomenon has occurred more recently? Would you consider the change in employment in this sector to represent a success or a failure from the standpoint of society as a whole?
- 5 Suppose that society decided to reduce consumption and increase investment.
 - a. How would this change affect economic growth?
 - b. What groups in society would benefit from this change? What groups might be hurt?
- 6 Societies choose what share of their resources to devote to consumption and what share to devote to investment. Some of these decisions involve private spending; others involve government spending.
 - a. Describe some forms of private spending that represent consumption, and some forms that represent investment.
 - b. Describe some forms of government spending that represent consumption, and some forms that represent investment.
- 7 What is the opportunity cost of investing in capital? Do you think a country can 'over-invest' in capital? What is the opportunity cost of investing in human capital? Do you think a country can 'over-invest' in human capital? Explain.
- 8 Suppose that a car company owned entirely by South Korean citizens opens a new factory in the north of England.
 - a. What sort of foreign investment would this represent?
 - b. What would be the effect of this investment on UK GDP? Would the effect on UK GNP be larger or smaller?
- 9 In the countries of South Asia in 1992, only 56 young women were enrolled in secondary school for every 100 young men. Describe several ways in which greater educational opportunities for young women could lead to faster economic growth in these countries.
- 10 International data show a positive correlation between political stability and economic growth.
 - a. Through what mechanism could political stability lead to strong economic growth?
 - b. Through what mechanism could strong economic growth lead to political stability?

23

UNEMPLOYMENT

Losing a job can be the most distressing economic event in a person's life. Most people rely on their labour earnings to maintain their standard of living, and many people get from their work not only income but also a sense of personal accomplishment. A job loss means a lower living standard in the present, anxiety about the future and reduced self-esteem. It is not surprising, therefore, that politicians campaigning for office often speak about how their proposed policies will help create jobs.

A major determinant of a country's standard of living is the amount of unemployment it typically experiences. People who would like to work but cannot find a job are not contributing to the economy's production of goods and services. Although some degree of unemployment is inevitable in a complex economy with thousands of firms and millions of workers, the amount of unemployment varies substantially over time and across countries. When a country keeps its workers as fully employed as possible, it achieves a higher level of GDP than it would if it left many of its workers standing idle.

This chapter begins our study of unemployment. The problem of unemployment is usefully divided into two categories – the long-run problem and the short-run problem. In this chapter we discuss the determinants of the long-run problem – an economy's *natural rate of unemployment*. As we will see, the designation *natural* does not imply that this rate of unemployment is desirable. Nor does it imply that it is constant over time or impervious to economic policy. It merely means that this unemployment does not go away on its own, even in the long run.

We begin the chapter by looking at three questions: how does the government measure the economy's rate of unemployment? What problems arise in interpreting the unemployment data? How long are the unemployed typically without work?

We then turn to the reasons why economies always experience some unemployment and the ways in which policymakers can help the unemployed. We discuss four explanations for the economy's natural rate of unemployment: job search, minimum wage laws, unions and efficiency wages. As we will see, long-run unemployment does not arise from a single problem that has a single solution. Instead, it reflects a variety of related problems. As a result, there is no easy way for policymakers to reduce the economy's natural rate of unemployment and, at the same time, to alleviate the hardships experienced by the unemployed.

We finish the chapter by looking at some of the costs of unemployment.

IDENTIFYING UNEMPLOYMENT

We begin this chapter by examining more precisely what the term *unemployment* means. We consider how governments measure unemployment, what problems arise in interpreting the unemployment data, how long the typical spell of unemployment lasts and why there will always be some people unemployed.

What is Unemployment?

The answer to this question may seem obvious: an unemployed person is someone who does not have a job. But as economists we need to be precise and careful in our definitions of economic categories. If you are in full-time education, for example, you do not have a full-time job in the usual sense of the word, i.e. you are not in full-time paid employment. And there is a good reason: you are studying. Hence you are

not available for work. What if you were not a student but were suffering from some long-term illness that meant that you were unfit for work? Again, although you would not have a job, we would not say that you were unemployed because you would not be available for work. From these two examples, it seems clear that we need to qualify our original definition of an unemployed person as ‘someone who does not have a job’ to ‘someone who does not have a job and who is available for work’.

But we still need to be clear as to what we mean by ‘available for work’. Suppose you were not in full-time employment and were looking for a job and I offered you a job as my research assistant for €1 a day. Would you take it? If we ignore for a moment the complication that economic research is so interesting that it is its own reward, you would probably not take the job because the wage rate offered is so low. At another extreme, suppose you won so much money on the Euro Millions Lottery that you decided you would leave university and live off your winnings for the rest of your life. Would you be unemployed? No, because you would still be unavailable for work, no matter what wage rate you were offered. Thus, being unemployed also depends upon whether you are willing to work (whether you are ‘available for work’) at going wage rates.

We are now in a position to give a more precise definition of what it means to be unemployed: the number unemployed in an economy is the number of people of working age who are able and available for work at current wage rates and who do not have a job.

Normally, economists find it more convenient to speak of the *unemployment rate*. This expresses the number unemployed as a percentage of the *labour force*, which in turn can be defined as the total number of people who could possibly be employed in the economy at any given point in time. If you think about it, this must be equal to the total number of people who are employed plus the total number of people who are unemployed.

How is Unemployment Measured?

How do government agencies go about measuring the unemployment rate in the economy? There are two basic ways.

The Claimant Count One simple way is to count the number of people who, on any given day, are claiming unemployment benefit payments from the government – the so-called *claimant count*. Since a government agency is paying out the benefits, it will be easy to gather data on the number of claimants. The government also has a good idea of the total labour force in employment, since it is receiving income tax payments from them. Adding to this the number of unemployment benefit claimants is a measure of the total labour force, and expressing the claimant count as a proportion of the labour force is a measure of the unemployment rate.

Since the government already has all the data necessary to compute the unemployment rate based on the claimant count, it is relatively cheap and easy to do. Unfortunately, there are a number of important drawbacks with the claimant count method.

One obvious problem is that it is subject to changes in the rules the government applies for eligibility to unemployment benefit. Suppose the government gets tougher and changes the rules so that fewer people are now entitled to unemployment benefit. The claimant count will go down and so will the measured unemployment rate, even though there has been no change in the number of people with or without work! The opposite would happen if the government became more lenient and relaxed the rules so that more people became eligible.

As it happens, governments do often change the rules on unemployment benefit eligibility. In the UK, for example, there have been about 30 changes to the eligibility rules over the past 25 years, all but one of which have reduced the claimant count and so reduced the unemployment rate based on this measure. The following are examples of categories of people who are excluded from the UK claimant count: people over the age of 55 who are without a job; those on government training programmes (largely school-leavers who have not found a job); anyone looking for part-time work; and people who have left the workforce for a while and now wish to return to employment (for example, women who have raised a family). Many – if not all – of the people in these categories would be people who do not have a job, are of working age and are able and available for work at current wage rates; yet they would be excluded from measured unemployment in the UK using the claimant count method.

Labour Force Surveys The second, and probably more reliable method of measuring unemployment is through the use of surveys – in other words, going out and asking people questions – based on an accepted definition of unemployment. Questions then arise as to whom to speak to, how often (since surveys use up resources and are costly) and what definition of unemployment to use. Although the definition of unemployment that we developed earlier seems reasonable enough, the term ‘available for work at current wage rates’ may be too loose for this purpose. In the UK and many other countries, the government carries out Labour Force Surveys (LFS) based on the standardized definition of unemployment from the International Labour Office, or ILO. The ILO definition of an unemployed person is someone who is without a job and who is willing to start work within the next two weeks and either has been looking for work within the past four weeks or was waiting to start a job.

The Labour Force Survey is carried out quarterly throughout Europe. National statistical services collect the data which are then processed by Eurostat. The surveys are published in different languages but scrutinized by statisticians to ensure comparability between the surveys carried out in each member state. In the UK, the survey is based on a sample of about 60,000 households and across Europe as a whole the sample size is around 1.5 million. Based on the answers to survey questions, the government places each adult (aged 16 and older) in each surveyed household into one of three categories:

- employed
- unemployed
- not in the labour force (or ‘economically inactive’).

A person is considered employed if he or she spent some of the previous week working at a paid job. A person is unemployed if he or she fits the ILO definition of an unemployed person. A person who fits neither of the first two categories, such as a full-time student, homemaker or retiree, is not in the labour force (or, to use ILO terminology, is economically inactive).

Once the government has placed all the individuals covered by the survey in a category, it computes various statistics to summarize the state of the labour market. The **labour force** is defined as the sum of the employed and the unemployed:

labour force the total number of workers, including both the employed and the unemployed

$$\text{Labour force} = \text{Number of employed} + \text{Number of unemployed}$$

Then the **unemployment rate** can be measured as the percentage of the labour force that is unemployed:

unemployment rate the percentage of the labour force that is unemployed:

$$\text{Unemployment rate} = \left(\frac{\text{Number of unemployed}}{\text{Labour force}} \right) \times 100$$

Unemployment rates are computed for the entire adult population and for more narrowly defined groups – men, women, youths and so on.

The same survey results are used to produce data on labour force participation. The **labour force participation rate** measures the percentage of the total adult population of the country that is in the labour force:

$$\text{Labour force participation rate} = \left(\frac{\text{Labour force}}{\text{Adult population}} \right) \times 100$$

labour force participation rate (or economic activity rate) the percentage of the adult population that is in the labour force

This statistic tells us the fraction of the population that has chosen to participate in the labour market. The labour force participation rate, like the unemployment rate, is computed both for the entire adult population and for more specific groups.

To see how these data are computed, consider the UK figures for January 2013. According to the *Labour Market Statistics*, 29.68 million people were employed and 2.49 million people were unemployed. The labour force was:

$$\text{Labour force} = 29.68 + 2.49 = 32.17 \text{ million}$$

The unemployment rate was:

$$\text{Unemployment rate} = \left(\frac{2.49}{32.17} \right) \times 100 = 7.74 \text{ per cent}$$

Because the adult population (the number of people aged between 15 and 65) was 41.7 million, the labour force participation rate was:

$$\text{Labour force participation rate} = (32.17/41.7) \times 100 = 77.14 \text{ per cent}$$

Hence, in January 2013, just over three-quarters of the UK adult population were participating in the labour market, and 7.74 per cent of those labour market participants were without work.

Table 23.1 shows the unemployment rate across the EU.

TABLE 23.1**Unemployment Rates in European Union (November 2012)**

| Country | Unemployment rate (%) |
|-----------------|-----------------------|
| Austria | 4.5 |
| Belgium | 7.4 |
| Bulgaria | 12.4 |
| Cyprus | 14.0 |
| Czech Republic | 7.4 |
| Denmark | 7.9 |
| Estonia | 9.5 |
| Finland | 7.9 |
| France | 10.5 |
| Germany | 5.4 |
| Greece | 26.0 |
| Hungary | 10.9 |
| Ireland | 14.6 |
| Italy | 11.1 |
| Latvia | 14.1 |
| Lithuania | 12.5 |
| Luxembourg | 5.1 |
| Malta | 6.9 |
| Netherlands | 5.6 |
| Poland | 10.6 |
| Portugal | 16.3 |
| Romania | 6.7 |
| Slovak Republic | 14.5 |
| Slovenia | 9.6 |
| Spain | 26.6 |
| Sweden | 8.1 |

The Natural Rate of Unemployment

Data on the labour market also allow economists and policymakers to monitor changes in the economy over time. Figure 23.1 shows the unemployment rate in the UK since 1971. The figure shows that the economy always has some unemployment and that the amount changes – often considerably – from year to year.

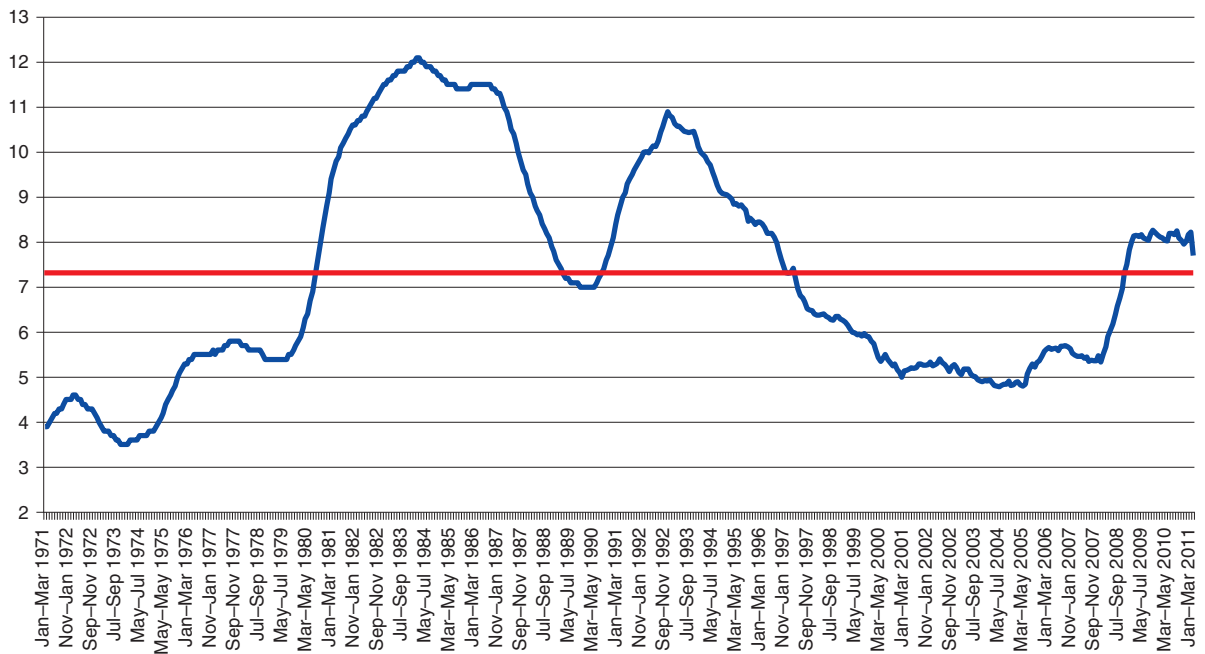
The normal rate of unemployment, around which the unemployment rate fluctuates is called the **natural rate of unemployment** and the deviation of unemployment from its natural rate is called **cyclical unemployment**. Figure 23.1 shows that the economy always has some unemployment; a trend line has been added to give some indication of the natural rate of unemployment (around 7.2 per cent).

natural rate of unemployment the normal rate of unemployment around which the unemployment rate fluctuates
cyclical unemployment the deviation of unemployment from its natural rate

FIGURE 23.1

UK Unemployment Rate Since 1971

This graph uses annual data on the unemployment rate to show the fraction of the labour force without a job, calculated by the LFS definition. A trend line has been added to show the natural rate of unemployment, the normal level of unemployment around which the unemployment rate fluctuates.



Later in this book we discuss short-run economic fluctuations, including the year-to-year fluctuations in unemployment. In the rest of this chapter, however, we ignore the short-run fluctuations and examine why there is always some unemployment in market economies.

How Long Are the Unemployed without Work?

In judging how serious the problem of unemployment is, one question to consider is whether unemployment is typically a short-term or long-term condition. If unemployment is short term, one might conclude that it is not a big problem. Workers may require a few weeks between jobs to find the openings that best suit their tastes and skills. Yet if unemployment is long term, one might conclude that it is a serious problem. Workers unemployed for many months are more likely to suffer economic and psychological hardship.

CASE STUDY Long-Term Unemployment in the European Union

A report published by the European Employment Observatory Review (see European Employment Observatory Review *Long Term Unemployment 2012. September 2012*) highlights some key trends in long-term unemployment in the EU. The report notes the significant costs both to the individual and to society of long-term unemployment which include loss of self-esteem, increasing difficulties of finding work because of the erosion of skills, damage to health, material deprivation, social exclusion and the cost of supporting the long-term unemployed.

Some 10 million people in the EU have been unemployed for longer than a year. Of these, 1.9 million have not had a job for over two years, 3 million have not had jobs for up to 17 months and 3.2 million have been out of work for between 24 and 47 months. The overall rate of long-term unemployment (LTU) in the EU sits at around 4.1 per cent in 2011 but some countries are faring better than others. In Spain the LTU rate has risen from 2 per cent in 2008 to 9 per cent in 2011 whereas in Germany the rate has decreased from 4.9 per cent in 2007 to 2.8 per cent in 2011.

The countries which have the lowest rates of LTU tend to be the more developed countries of northern Europe such as Norway, Sweden, Luxembourg, the Netherlands, Austria and Denmark, all of whom have LTU rates under 2.0 per cent, whereas the former Soviet Bloc countries have the highest: Latvia, Hungary, Estonia and Slovakia all have rates in excess of 8.0 per cent. Ireland too has a LTU rate of over 8.0 per cent, largely due to the effects of the financial crisis in 2007–2008 which led to the rate increasing from under 2.0 per cent to around 8.5 per cent.

One of the dangers of long-term unemployment is that workers find it impossible to find work because the longer they are out of work the less likely it is that firms will see these people as being serious contenders for jobs. The result is what is called **hysteresis** – the lagging effects of past economic events on future ones. Those who are unemployed for long periods gradually adjust to a lower standard of living, they also find it increasingly harder to get work and so may lose interest in returning to the labour market.

hysteresis the lagging effects of past economic events on future ones

When looking at the breakdown of the LTU, it seems that older people are more likely to be victims of LTU but a worrying trend is the increasing number of young people across the EU who are falling into the LTU bracket. Prior to the financial crisis it was women who were more likely to be LTU but this has now shifted and more men are now LTU.



The rise in youth unemployment across Europe is a grave concern. For many young people without jobs, the future looks bleak.

Because the duration of unemployment can affect our view about how big a problem it is, economists have devoted much energy to studying data on the duration of unemployment spells. In this work, they have uncovered a result that is important, subtle and seemingly contradictory: most spells of unemployment are short, and most unemployment observed at any given time is long term.

To see how this statement can be true, consider an example. Suppose that you visited the government's unemployment office every week for a year to survey the unemployed. Each week you find that there are four unemployed workers. Three of these workers are the same individuals for the whole year, while the fourth person changes every week. Based on this experience, would you say that unemployment is typically short term or long term?

Some simple calculations help answer this question. In this example, you meet a total of 55 unemployed people: 52 of them are unemployed for one week, and three are unemployed for the full year. This means that 52/55, or 95 per cent, of unemployment spells end in one week. Thus, most spells of unemployment are short. Yet consider the total amount of unemployment. The three people unemployed for one year (52 weeks) make up a total of 156 weeks of unemployment. Together with the 52 people unemployed for one week, this makes 208 weeks of unemployment. In this example, 156/208, or 75 per cent, of unemployment is attributable to those individuals who are unemployed for a full year. Thus, most unemployment observed at any given time is long term.

This subtle conclusion implies that economists and policymakers must be careful when interpreting data on unemployment and when designing policies to help the unemployed. Most people who become unemployed will soon find jobs. Yet most of the economy's unemployment problem is attributable to the relatively few workers who are jobless for long periods of time.

Why are There Always Some People Unemployed?

In most markets in the economy, prices adjust to bring quantity supplied and quantity demanded into balance. In an ideal labour market, wages would adjust to balance the quantity of labour supplied and the quantity of labour demanded. This adjustment of wages would ensure that all workers are always fully employed.

Of course, reality does not resemble this ideal. There are always some workers without jobs, even when the overall economy is doing well. In other words, the unemployment rate never falls to zero; instead, it fluctuates around the natural rate of unemployment. To understand this natural rate, we will examine the reasons why actual labour markets depart from the ideal of full employment.

To preview our conclusions, we will find that there are four ways to explain unemployment in the long run.

The first explanation is that it takes time for workers to search for the jobs that are best suited for them. The unemployment that results from the process of matching workers and jobs is sometimes called **frictional unemployment**, and it is often thought to explain relatively short spells of unemployment.

frictional unemployment unemployment that results because it takes time for workers to search for the jobs that best suit their tastes and skills

The next three explanations for unemployment suggest that the number of jobs available in some labour markets may be insufficient to give a job to everyone who wants one. This occurs when the quantity of labour supplied exceeds the quantity demanded. Unemployment of this sort is sometimes called **structural unemployment**, and it is often thought to explain longer spells of unemployment. As we will see, this kind of unemployment results when wages are, for some reason, set above the level that brings supply and demand into equilibrium. We will examine three possible reasons for an above-equilibrium wage: minimum wage laws, unions and efficiency wages.

structural unemployment unemployment that results because the number of jobs available in some labour markets is insufficient to provide a job for everyone who wants one

SELF TEST How is the unemployment rate measured? • How might the unemployment rate overstate the amount of joblessness? How might it understate it?

In Chapter 31 we will introduce the ideas of John Maynard Keynes whose ideas on unemployment changed the nature of economics. Since the 1930s there has been an ongoing debate in economics about the principal causes of unemployment and the extent to which these causes stem from the supply side of the economy or the demand side. That debate has continued and has been brought into sharper focus with the rise in unemployment in many European countries during the first decade of the new century. Policies designed to improve the workings of the supply side of the economy which includes cutting taxes, investing in training and education, changing the benefits system and improving incentives to get work, have to be viewed alongside those which affect the level of aggregate demand in the economy. In essence, the debate centres on the extent to which unemployment is cyclical or structural – the latter accounting for the level of unemployment not due to changes in the economic cycle. As we build the theoretical building blocks of macroeconomics it must be remembered that policies to cut unemployment still deeply divide opinion in the economics profession.

JOB SEARCH

One reason why economies always experience some unemployment is job search. **Job search** is the process of matching workers with appropriate jobs. If all workers and all jobs were the same, so that all workers were equally well suited for all jobs, job search would not be a problem. Laid-off workers would quickly find new jobs that were well suited for them. But, in fact, workers differ in their tastes and skills, jobs differ in their attributes, and information about job candidates and job vacancies is disseminated slowly among the many firms and households in the economy.

job search the process by which workers find appropriate jobs given their tastes and skills

Why Some Frictional Unemployment Is Inevitable

Frictional unemployment is often the result of changes in the demand for labour among different firms. When consumers decide that they prefer Brand **X** to Brand **Y**, the company producing Brand **X** increases employment, and the other firm lays off workers. The former Brand **Y** workers must now search for new jobs, and the Brand **X** producer must decide which new workers to hire for the various jobs that have opened up. The result of this transition is a period of unemployment.

Similarly, because different regions of the country produce different goods, employment can rise in one region while it falls in another. Consider, for instance, what happens when the world price of oil falls. Firms extracting oil from the fields below the North Sea, off the coast of Scotland, respond to the lower price by cutting back on production and employment. At the same time, cheaper petrol stimulates car sales, so car manufacturing firms in northern and central England raise production and employment. Changes in the composition of demand among industries or regions are called *sectoral shifts*. Because it takes time for workers to search for jobs in the new sectors, sectoral shifts temporarily cause unemployment.

Frictional unemployment is inevitable simply because the economy is always changing. In 1960, manufacturing as a percentage of GDP was 38 per cent in the UK; it now accounts for around 15 per cent. The number of people employed in manufacturing in 1960 was around 9 million compared to about 3 million today. The gross value added by industry (where gross value added is defined as the value of newly generated goods and services) in the EU 27 (which includes manufacturing) fell from 23.1 per cent in 1998 to 18.7 per cent in 2010. Even Germany, traditionally heavily reliant on industry for its export earnings, has seen the value added by industry fall from 25.0 per cent in 2001 to 23.8 per cent in 2010.

On the other hand, business services and finance contributed only about 3 per cent of UK GDP in the mid-1950s but contribute more than a quarter today. The EU 27 has seen gross value added by business activities and financial services increasing from 23.1 per cent of GDP in 1998 to 29.0 per cent in 2010. As these transitions take place, jobs are created in some firms and destroyed in others. The end result of this process has been higher productivity and higher living standards. But, along the way, workers in declining industries found themselves out of work and searching for new jobs.

In addition to the effects of sectoral shifts on unemployment, workers will leave their jobs sometimes because they realize that the jobs are not a good match for their tastes and skills and they wish to look for a better job. Many of these workers, especially younger ones, find new jobs at higher wages, although given the vast improvements in information technology in recent years (especially the Internet) it is likely that many people search for new jobs without actually quitting their current job. Nevertheless, this churning of the labour force is normal in a well-functioning and dynamic market economy, and the result is some amount of frictional unemployment.

Public Policy and Job Search

Even if some frictional unemployment is inevitable, the precise amount is not. The faster information spreads about job openings and worker availability, the more rapidly the economy can match workers and firms. The Internet, for instance, may help facilitate job search and reduce frictional unemployment. In addition, public policy may play a role. If policy can reduce the time it takes unemployed workers to find new jobs, it can reduce the economy's natural rate of unemployment.

Government policies try to facilitate job search in various ways. One way is through government-run employment agencies or job centres, which give out information about job vacancies. Another way is through public training schemes, which aim to ease the transition of workers from declining to growing industries and to help disadvantaged groups escape poverty. Advocates of these policies believe that they make the economy operate more efficiently by keeping the labour force more fully employed, and that they reduce the inequities inherent in a constantly changing market economy.

Critics of these policies question whether the government should get involved with the process of job search. They argue that it is better to let the private market match workers and jobs. In fact, most job search in the economy takes place without intervention by the government. Newspaper advertisements, internet job sites, head-hunters and word of mouth all help spread information about job openings and job candidates. Similarly, much worker education is done privately, either through schools or through on-the-job training. These critics contend that the government is no better – and most likely worse – at disseminating the right information to the right workers and deciding what kinds of worker training would be most valuable. They claim that these decisions are best made privately by workers and employers.

Unemployment Insurance

One government policy that increases the amount of frictional unemployment, without intending to do so, is **unemployment insurance** (or, as it is called in the UK, national insurance). This policy is designed to offer workers partial protection against job loss. The unemployed who quit their jobs, were fired for just cause or who have just entered the labour force are not eligible. Benefits are paid only to the unemployed who were laid off because their previous employers no longer needed their skills.

unemployment insurance a government programme that partially protects workers' incomes when they become unemployed

While unemployment insurance reduces the hardship of unemployment, it is argued that it can also increase the amount of unemployment. This explanation is based on one of the *Ten Principles of Economics*: people respond to incentives. Because unemployment benefits stop when a worker takes a new job, the unemployed, it is argued, devote less effort to job search and are more likely to turn down unattractive

job offers. In addition, because unemployment insurance makes unemployment less onerous, workers are less likely to seek guarantees of job security when they negotiate with employers over the terms of employment. However, research on unemployment insurance in Europe gives a different perspective. In a paper by Konstantinos Tatsiramos (Tatsiramos, K. (2006) *Unemployment Insurance in Europe: Unemployment Duration and Subsequent Employment Stability*. Institute for the Study of Labor Discussion Paper no. 2280) the benefits to workers searching for jobs and receiving unemployment insurance is greater than the costs:

This paper provides evidence on the effect of unemployment benefits on unemployment and employment duration in Europe, using individual data from the European Community Household Panel for eight countries. Even if receiving benefits has a direct negative effect increasing the duration of unemployment spells, there is also a positive indirect effect of benefits on subsequent employment duration. This indirect effect is pronounced in countries with relatively generous benefit systems, and for recipients who have remained unemployed for at least six months. In terms of the magnitude of the effect, recipients remain employed on average two to four months longer than non-recipients. This represents a 10 to 20 per cent increase relative to the average employment duration, compensating for the additional time spent in unemployment.

The effect of unemployment insurance is likely to be related to the way the scheme is designed and operated. In one US study, when unemployed workers applied to collect unemployment insurance benefits, some of them were randomly selected and offered each a \$500 bonus if they found new jobs within 11 weeks. This group was then compared with a control group not offered the incentive. The average spell of unemployment for the group offered the bonus was 7 per cent shorter than the average spell for the control group. This experiment shows that the design of the unemployment insurance system influences the effort that the unemployed devote to job search.

Several other studies examined search effort by following a group of workers over time. Unemployment insurance benefits, rather than lasting forever, usually run out after six months or a year. These studies found that when the unemployed become ineligible for benefits, the probability of their finding a new job rises markedly. Thus, receiving unemployment insurance benefits does reduce the search effort of the unemployed.

Even though unemployment insurance reduces search effort and raises unemployment, we should not necessarily conclude that the policy is a bad one. The policy does achieve its primary goal of reducing the income uncertainty that workers face. In addition, when workers turn down unattractive job offers, they have the opportunity to look for jobs that better suit their tastes and skills. Some economists have argued that unemployment insurance improves the ability of the economy to match each worker with the most appropriate job.

The study of unemployment insurance shows that the unemployment rate is an imperfect measure of a nation's overall level of economic well-being. Most economists agree that eliminating unemployment insurance would reduce the amount of unemployment in the economy. Yet economists disagree on whether economic well-being would be enhanced or diminished by this change in policy.

SELF TEST How would an increase in the world price of oil affect the amount of frictional unemployment? Is this unemployment undesirable? What public policies might affect the amount of unemployment caused by this price change?

MINIMUM WAGE LAWS

Having seen how frictional unemployment results from the process of matching workers and jobs, let's now examine how structural unemployment results when the number of jobs is insufficient for the number of workers.

To understand structural unemployment, we begin by reviewing how unemployment arises from minimum wage laws. Although minimum wages are not the predominant reason for unemployment in an

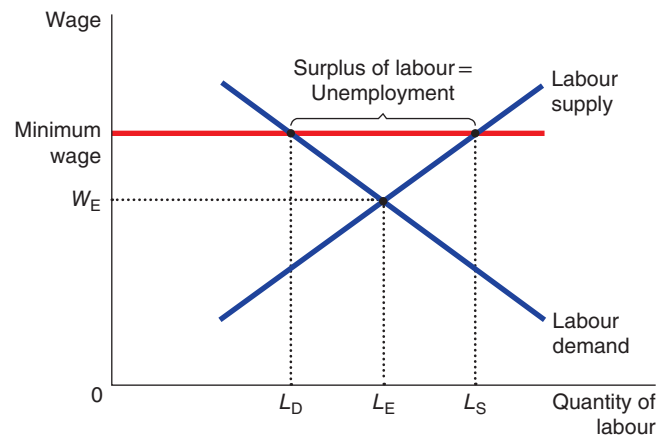
economy, they have an important effect on certain groups with particularly high unemployment rates. Moreover, the analysis of minimum wages can be used to understand some of the other reasons for structural unemployment.

Figure 23.2 reviews the basic economics of a minimum wage. When a minimum wage law forces the wage to remain above the level that balances supply and demand, it raises the quantity of labour supplied and reduces the quantity of labour demanded compared to the equilibrium level. There is a surplus of labour. Because there are more workers willing to work than there are jobs, some workers are unemployed.

FIGURE 23.2

Unemployment From a Wage Above the Equilibrium Level

In this labour market, the wage at which supply and demand balance is W_E . At this equilibrium wage, the quantity of labour supplied and the quantity of labour demanded both equal L_E . By contrast, if the wage is forced to remain above the equilibrium level, perhaps because of a minimum wage law, the quantity of labour supplied rises to L_S , and the quantity of labour demanded falls to L_D . The resulting surplus of labour, $L_S - L_D$, represents unemployment.



Because we discussed minimum wage laws in Chapter 17, we will not discuss them further here. It is, however, important to note why minimum wage laws are not a predominant reason for unemployment: most workers in the economy have wages well above the legal minimum. Minimum wage laws are binding most often for the least skilled and least experienced members of the labour force, such as teenagers. It is only among these workers that minimum wage laws explain the existence of unemployment.

Although Figure 23.2 is drawn to show the effects of a minimum wage law, it also illustrates a more general lesson: if the wage is kept above the equilibrium level for any reason, the result is unemployment. Minimum wage laws are just one reason why wages may be 'too high'. In the next two sections of this chapter, we consider two other reasons why wages may be kept above the equilibrium level – unions and efficiency wages. The basic economics of unemployment in these cases is the same as that shown in Figure 23.2, but these explanations of unemployment can apply to many more of the economy's workers.

At this point, however, we should stop and notice that the structural unemployment that arises from an above-equilibrium wage is, in an important sense, different from the frictional unemployment that arises from the process of job search. The need for job search is not due to the failure of wages to balance labour supply and labour demand. When job search is the explanation for unemployment, workers are *searching* for the jobs that best suit their tastes and skills. By contrast, when the wage is above the equilibrium level, the quantity of labour supplied exceeds the quantity of labour demanded, and workers are unemployed because they are *waiting* for jobs to open up.

SELF TEST Draw the supply curve and the demand curve for a labour market in which the wage is fixed above the equilibrium level. Show the quantity of labour supplied, the quantity demanded and the amount of unemployment.

UNIONS AND COLLECTIVE BARGAINING

A union is a worker association that bargains with employers over wages and working conditions. **Union density** measures the proportion of the workforce that is unionized, excluding people who cannot, for legal or other reasons, be members of a union – for example, members of the armed forces.

union density a measure of the proportion of the workforce that is unionized

Broadly speaking, this amounts to expressing the number of union members as a proportion of civilian employees plus the unemployed. In the UK in 2011, union density was 23.8 per cent, and has been steadily falling since 1995 when it stood at around 32.4 per cent and an even greater marked fall from the beginning of the 1980s, when it was over 50 per cent. In other European countries there is a similar trend of falling union density. In Germany density has fallen from 23.3 per cent in 1999 to 18.5 per cent in 2010, in the Netherlands the fall has been from 24.6 per cent to 18.2 per cent in 2011. However, there are exceptions with countries like Finland, Denmark, Iceland and Sweden having densities between 68 per cent and 79 per cent, however even in these countries the density is gradually falling. In the United States, by comparison, density was around 11.3 per cent and in Australia 18.0 per cent (both in 2011, data from the OECD).

The Economics of Unions

A union is a type of cartel. Like any cartel, a union is a group of sellers acting together in the hope of exerting their joint market power. Workers in a union act as a group when discussing their wages, benefits and working conditions with their employers. The process by which unions and firms agree on the terms of employment is called **collective bargaining**.

collective bargaining the process by which unions and firms agree on the terms of employment

When a union bargains with a firm, it asks for higher wages, better benefits and better working conditions than the firm would offer in the absence of a union. If the union and the firm do not reach agreement, the union can take various steps to put pressure on employers to come to an agreement including working to rule (doing only what is agreed in the contract of employment) and as a last resort organizing a withdrawal of labour from the firm, called a strike. Because a strike reduces production, sales and profit, a firm facing a strike threat is likely to agree to pay higher wages than it otherwise would. Economists who study the effects of unions typically find that union workers earn significantly more than similar workers who do not belong to unions.

When a union raises the wage above the equilibrium level, it raises the quantity of labour supplied and reduces the quantity of labour demanded, resulting in unemployment. Those workers who remain employed are better off, but those who were previously employed and are now unemployed are worse off. Indeed, unions are often thought to cause conflict between different groups of workers – between the *insiders* who benefit from high union wages and the *outsiders* who do not get the union jobs.

The outsiders can respond to their status in one of two ways. Some of them remain unemployed and wait for the chance to become insiders and earn the high union wage. Others take jobs in firms that are

not unionized. Thus, when unions raise wages in one part of the economy, the supply of labour increases in other parts of the economy. This increase in labour supply, in turn, reduces wages in industries that are not unionized. In other words, workers in unions reap the benefit of collective bargaining, while workers not in unions bear some of the cost.

The role of unions in the economy depends in part on the laws that govern union organization and collective bargaining. Normally, explicit agreements among members of a cartel are illegal. If firms that sell a common product were to agree to set a high price for that product, they would generally be held to be in breach of competition law and the government would prosecute these firms in the civil and criminal courts. In contrast, unions are given exemption from these laws in the belief that workers need greater market power as they bargain with employers.

Legislation affecting the market power of unions is a perennial topic of political debate. Members of parliament sometimes debate *right-to-work laws*, which give workers in a unionized firm the right to choose whether to join the union. In the absence of such laws, unions can insist during collective bargaining that firms make union membership a requirement for employment.

Are Unions Good or Bad for the Economy?

Economists disagree about whether unions are good or bad for the economy as a whole. Let's consider both sides of the debate.

Critics of unions argue that unions are merely a type of cartel. When unions raise wages above the level that would prevail in competitive markets, they reduce the quantity of labour demanded, cause some workers to be unemployed and reduce the wages in the rest of the economy. The resulting allocation of labour is, critics argue, both inefficient and inequitable. It is inefficient because high union wages reduce employment in unionized firms below the efficient, competitive level. It is inequitable because some workers benefit at the expense of other workers.

Advocates of unions contend that unions are a necessary antidote to the market power of the firms that hire workers. In some regions where one particular company is the dominant employer, if workers do not accept the wages and working conditions that the firm offers, they may have little choice but to move or stop working. In the absence of a union, therefore, the firm could use its market power to pay lower wages and offer worse working conditions than would prevail if it had to compete with other firms for the same workers. In this case, a union may balance the firm's market power and protect the workers from being at the mercy of the firm owners.

Advocates of unions also claim that unions are important for helping firms respond efficiently to workers' concerns. Whenever a worker takes a job, the worker and the firm must agree on many attributes of the job in addition to the wage: hours of work, overtime, holidays, sick leave, health benefits, promotion schedules, job security and so on. By representing workers' views on these issues, unions allow firms to provide the right mix of job attributes. In many countries unions have now taken on additional roles in supporting workers with respect to offering legal support in the event of an individual dispute at work, advice on pensions, financial services such as insurance, and support for those who have been injured or disabled at work and have to retire early. Even if unions have the adverse effect of pushing wages above the equilibrium level and causing unemployment, they have the benefit of helping firms keep a happy and productive workforce.

In the end, there is no consensus among economists about whether unions are good or bad for the economy. Like many institutions, their influence is probably beneficial in some circumstances and adverse in others.

SELF TEST How does a union in the car industry affect wages and employment at Ford and Nissan plants in the UK? How might it affect wages and employment in other industries?

THE THEORY OF EFFICIENCY WAGES

A fourth reason why economies always experience some unemployment is suggested by the theory of efficiency wages. According to this theory, firms operate more efficiently if wages are above the equilibrium level. Therefore, it may be profitable for firms to keep wages high even in the presence of a surplus of labour.

In some ways, the unemployment that arises from efficiency wages is similar to the unemployment that arises from minimum wage laws and unions. In all three cases, unemployment is the result of wages above the level that balances the quantity of labour supplied and the quantity of labour demanded. Yet there is also an important difference. Minimum wage laws and unions prevent firms from lowering wages in the presence of a surplus of workers. Efficiency wage theory states that such a constraint on firms is unnecessary in many cases because firms may be better off keeping wages above the equilibrium level.

Why should firms want to keep wages high? This decision may seem odd at first, for wages are a large part of firms' costs. Normally, we expect profit-maximizing firms to want to keep costs – and therefore wages – as low as possible. The novel insight of efficiency wage theory is that paying high wages might be profitable because they might raise the efficiency of a firm's workers.

There are several types of efficiency wage theory. Each type suggests a different explanation for why firms may want to pay high wages. Let's now consider four of these types.

Worker Health

The first and simplest type of efficiency wage theory emphasizes the link between wages and worker health. Better paid workers eat a more nutritious diet, and workers who eat a better diet are healthier and more productive. A firm may find it more profitable to pay high wages and have healthy, productive workers than to pay lower wages and have less healthy, less productive workers.

This type of efficiency wage theory is not relevant for firms in rich countries such as the United Kingdom and many of those in Europe. In these countries, the equilibrium wages for most workers are well above the level needed for an adequate diet. Firms are not concerned that paying equilibrium wages would place their workers' health in jeopardy.

This type of efficiency wage theory is more relevant for firms in less developed countries where inadequate nutrition is a more common problem. Unemployment is high in the cities of many poor African countries, for example. In these countries, firms may fear that cutting wages would, in fact, adversely influence their workers' health and productivity. In other words, concern over nutrition may explain why firms do not cut wages despite a surplus of labour.

Worker Turnover

A second type of efficiency wage theory emphasizes the link between wages and worker turnover. Workers quit jobs for many reasons – to take jobs in other firms, to move to other parts of the country, to leave the labour force and so on. The frequency with which they quit depends on the entire set of incentives they face, including the benefits of leaving and the benefits of staying. The more a firm pays its workers, the less often its workers will choose to leave. Thus, a firm can reduce turnover among its workers by paying them a high wage.

Why do firms care about turnover? The reason is that it is costly for firms to hire and train new workers. Moreover, even after they are trained, newly hired workers are not as productive as experienced workers. Firms with higher turnover, therefore, will tend to have higher production costs. Firms may find it profitable to pay wages above the equilibrium level in order to reduce worker turnover.

Worker Effort

A third type of efficiency wage theory emphasizes the link between wages and worker effort. In many jobs, workers have some discretion over how hard to work. As a result, firms monitor the efforts

of their workers, and workers caught shirking their responsibilities can be disciplined and possibly dismissed. But not all shirkers are caught immediately because monitoring workers is costly and imperfect. A firm can respond to this problem by paying wages above the equilibrium level. High wages make workers more eager to keep their jobs and, thereby, give workers an incentive to put forward their best effort.

This particular type of efficiency wage theory is similar to the old Marxist idea of the 'reserve army of the unemployed'. Marx thought that employers benefited from unemployment because the threat of unemployment helped to discipline those workers who had jobs. In the worker effort variant of efficiency wage theory, unemployment fills a similar role. If the wage were at the level that balanced supply and demand, workers would have less reason to work hard because if they were fired, they could quickly find new jobs at the same wage. Therefore, firms raise wages above the equilibrium level, causing unemployment and providing an incentive for workers not to shirk their responsibilities.

Worker Quality

A fourth and final type of efficiency wage theory emphasizes the link between wages and worker quality. When a firm hires new workers, it cannot perfectly gauge the quality of the applicants. By paying a high wage, the firm attracts a better pool of workers to apply for its jobs.

To see how this might work, consider a simple example. Waterwell Company owns one well and needs one worker to pump water from the well. Two workers, Jekyll and Hyde, are interested in the job. Jekyll, a proficient worker, is willing to work for €10 per hour. Below that wage, he would rather start his own car-washing business. Hyde, a complete incompetent, is willing to work for anything above €2 per hour. Below that wage, he would rather sit on the beach. Economists say that Jekyll's *reservation wage* – the lowest wage he would accept – is €10 per hour, and Hyde's reservation wage is €2 per hour.

What wage should the firm set? If the firm was interested in minimizing labour costs, it would set the wage at €2 per hour. At this wage, the quantity of workers supplied (one) would balance the quantity demanded. Hyde would take the job, and Jekyll would not apply for it. Yet suppose Waterwell knows that only one of these two applicants is competent, but it does not know whether it is Jekyll or Hyde. If the firm hires the incompetent worker, he may damage the well, causing the firm huge losses. In this case, the firm has a better strategy than paying the equilibrium wage of €2 per hour and hiring Hyde. It can offer €10 per hour, inducing both Jekyll and Hyde to apply for the job. By choosing randomly between these two applicants and turning the other away, the firm has a 50:50 chance of hiring the competent one. By contrast, if the firm offers any lower wage, it is sure to hire the incompetent worker.

This story illustrates a general phenomenon. When a firm faces a surplus of workers, it might seem profitable to reduce the wage it is offering. But by reducing the wage, the firm induces an adverse change in the mix of workers. In this case, at a wage of €10 per hour, Waterwell has two workers applying for one job. But if Waterwell responds to this labour surplus by reducing the wage, the competent worker (who has better alternative opportunities) will not apply. Thus, it is profitable for the firm to pay a wage above the level that balances supply and demand.

SELF TEST Give four explanations for why firms might find it profitable to pay wages above the level that balances quantity of labour supplied and quantity of labour demanded.

THE COST OF UNEMPLOYMENT

We mentioned at the outset of this chapter that unemployment can be one of the most distressing things to happen to anyone in their lives. Unemployment imposes costs on the individual and their family and friends but there is also a wider cost of unemployment which affects government, the taxpayer and the economy as a whole.

The Costs of Unemployment to the Individual

Loss of Earnings One of the first and perhaps most obvious costs of unemployment to an individual is the loss of earnings that results from being unemployed. Many countries provide some form of unemployment insurance as we have seen but the sums given to the unemployed are relatively small and in most cases nowhere near the earnings that the individual would have earned in work. In some countries there might be other state benefits that an unemployed worker can claim which can mean that in the case of workers in low-skilled, low-paid industries the incentive to take work can be reduced as we have seen. In many other cases, however, unemployment means that individuals have to re-evaluate their household spending budgets.

This might mean that households cut back on certain luxuries like spending on leisure activities, going to the cinema, going out to eat and so on, but might also cut down on luxuries such as clothing, furniture, electrical goods, having extensions built on a house and so on. We will see shortly how this has an effect on society as a whole.

The unemployed and their families are more likely to be at risk of slipping into poverty. Remember, the definition of poverty is a household income less than 60 per cent below the median income. It is unlikely that welfare support from the state is ever going to be, on its own, sufficient to put families above this level so unless the unemployed have savings to draw upon it is more likely that they will fall into poverty. The Trades Union Congress (TUC) in the UK estimate that around 60 per cent of working age adults in families where there is unemployment are likely to be in poverty.

In addition, some families will find that unemployment means that they face problems in paying for rent or mortgages and this can result in the loss of homes. The unemployed are more likely to have to go into debt in order to pay bills and this can add to the problems highlighted in our next section.

Stress, Self-Esteem and Health Problems Being unemployed can lead to significant mental health problems. The process of becoming unemployed is stressful, it can be a life-changing event for some people. Having to adjust to claiming benefits, applying for other jobs, possibly getting additional training and the chances of having repeated experiences of either not having any replies to applications or in the event of unsuccessful interviews, the feeling of rejection, is not only stressful in itself but can lead to feelings of guilt and the reduction in self-esteem. These experiences can bring on stress-related illnesses and the incidences of health problems in the unemployed can increase the longer that the unemployment continues.

Drug and Alcohol Abuse and Crime Closely linked to the self-esteem problems, the boredom that can result from being unemployed and the feeling of worthlessness that many unemployed people say they experience, is the increased potential to turn to alcohol and illegal drugs as a means of escape. In 2010, a survey by the Prince's Trust Youth Index reported that over 10 per cent of young people experiencing unemployment had abused drugs and alcohol.

When people move from being employed and having an income to be able to afford a reasonable standard of living, to experiencing tight restrictions on incomes and spending, the feeling of social exclusion and deprivation can be acute and cause some to turn to crime as a means of maintaining what they see as a reasonable standard of living. Indeed, the correlation between crime rates and drug abuse is very high; once people get involved with a drug or alcohol habit it becomes expensive and one way of funding this habit is to turn to crime.

Family Breakdown Families who have an experience of unemployment are more prone to break-up. Divorce rates amongst the unemployed are higher as the stresses of coping with adjusting to new income levels, trying to find work and so on take their toll.

De-Skilling The longer someone is out of work the more likely it is that they will lose touch with changes in the workplace and the labour market in general and the more likely it is that they might be viewed as being unemployable or not favourable candidates for employment. Changes in the workplace, in technology and in the skills needed for employment change rapidly. Those in work are able to take advantage of training – both off-the-job and on-the-job, to maintain their skill levels, but the unemployed may be

excluded from being able to maintain or improve their skill levels and as a result find it even harder to find work. This can lead to the hysteresis effect as outlined above.

The Costs of Unemployment to Society and the Economy

The Opportunity Cost of Unemployment An individual who is willing and able to work represents a unit of productive output. If that person is unemployed the opportunity cost to society is the value of the goods and services that the individual could have produced. This 'lost output' can be considerable and represents a lower standard of living for society as a whole. In Chapter 19 we looked at the production possibilities frontier (PPF). If there is unemployment in society which is not simply frictional unemployment then society will not be operating on the PPF but instead somewhere inside it which represents an inefficient use of resources.

The Tax and Benefits Effect People who are unemployed have lower incomes and may rely solely on government welfare payments to support their standard of living. If people lose their jobs then they do not pay as much in income taxes and if they also reduce spending they then do not pay consumption taxes at the same level as if they were in employment. The higher the level of unemployment in a country the greater the impact on tax revenue for the government. Not only is government revenue adversely affected but government spending is also likely to be higher. The unemployed will claim additional benefits and in addition, governments may also incur additional costs in having to deal with the social problems caused by unemployment such as drug and alcohol abuse, family breakdown and the increase in crime.

If government income is reduced through lower tax receipts and there is also a requirement to increase spending because of higher welfare spending, the pressure on government budgets can increase and it is more likely that the government experience a budget deficit – a situation where its spending is higher than its revenue from taxation. If governments experience a budget deficit then the deficit has to be funded by additional borrowing. If governments have to borrow more then this can not only cause crowding out but also puts upward pressure on interest rates, which in turn might affect investment decisions by firms. The knock-on effects can have ripples throughout the economy as our next point highlights.

The Reverse Multiplier Effect We have seen that when people experience unemployment they cut back their spending on luxuries and may also switch their spending to substitute goods which may be seen as inferior goods. Firms who produce these different goods may see a change in spending patterns which can have an effect on cash flows and ultimately profits.

Goods with a relatively high income elasticity of demand are likely to be affected more significantly. If sales fall firms earn lower incomes and may have to adjust their business to manage cash flows. This might involve cutting back on orders from suppliers, building up stocks as goods remain unsold and in some cases firms may have to either lay-off workers or even make workers redundant or close down the business if it becomes insolvent. If workers are made redundant or lose their jobs in this way this then means those workers receive lower incomes and so the process continues.

Not all firms will be affected in such an extreme way but it is the case that in periods of high unemployment, firms may see falling profit levels and this in turn means they pay lower corporate taxes which puts further pressure on government budgets. Some firms might see demand for their services actually increase in periods of high unemployment. If the unemployed switch spending to inferior goods, the producers of those goods might see demand rise. In the aftermath of the financial crisis and beyond, low-cost supermarkets across the UK and Europe reported seeing an increase in sales whilst traditional supermarkets reported reduced sales.

The effect of unemployment if more than simply frictional unemployment is to produce a multiplied impact on economic activity as a whole. If there are concentrated pockets of workers losing their jobs such as a major employer in an area, then the effect of this reverse multiplier effect can be considerable. Indeed, there are areas of the UK and Europe where the decline in industries concentrated in particular areas has led to considerable regional deprivation which has lasted for many years. Once an area is caught in the cycle of economic decline it is extremely hard to recover.

CONCLUSION

In this chapter we discussed the measurement of unemployment, the reasons why economies always experience some degree of unemployment and some of the costs of unemployment to society and the economy as a whole. We have seen how job search, minimum wage laws, unions and efficiency wages can all help explain why some workers do not have jobs. Which of these four explanations for the natural rate of unemployment are the most important? Unfortunately, there is no easy way to tell. Economists differ in which of these explanations of unemployment they consider most important.

The analysis of this chapter yields an important lesson: although the economy will always have some unemployment, its natural rate is not immutable. Many events and policies can change the amount of unemployment the economy typically experiences. As the information revolution changes the process of job search, as government adjusts the minimum wage, as workers form or quit unions, and as firms alter their reliance on efficiency wages, the natural rate of unemployment evolves. Unemployment is not a simple problem with a simple solution. But how we choose to organize our society can profoundly influence how prevalent a problem it is. The costs of unemployment are significant not only to individuals but also to the economy and society which is why so many governments around the world are highly sensitive to the problem and look to develop policies to deal with unemployment. It is worth noting that when high unemployment persists, the potential for social disorder and political upheaval is greater. It might not be surprising, therefore, that governments are keen to ensure that the problems which can arise from unemployment are given due consideration in policy formulation.

IN THE NEWS



Unemployment and Vacancies

You may sometimes hear news reports of the number of vacancies in the economy and wonder why unemployment is so high when there are seemingly a large number of jobs available. This article looks at this in more detail.

Search Costs in the Labour Market

According to the Organization of Economic Co operation and Development (OECD), in December 2012 there were around 28,045 job vacancies in Austria against a registered unemployment level (the claimant count) of 273,615 (10.2 per cent of the unemployed), in the UK, 494,000 vacancies against a claimant count of 1,557,100 (31.7 per cent), in Switzerland 16,866 against 131,509 (12.8 per cent) and in Sweden 45,132 against 217,187 (20.7 per cent). In theory, if the unemployed could fill the vacancies that exist then unemployment would

be lower and the costs to individuals and society would be less.

The relationship between vacancies and unemployment does tend to be regional in many countries. In some parts of the UK, the number of people chasing each job vacancy is high with figures from a job search website, Adzuna, suggesting around 80 applicants for every vacancy in some cities such as Hull in the north of England, 73 per job in Stoke-on-Trent and 53 in Sunderland, while in parts of the south east, such as Reading, Cambridge, London and Milton Keynes, the number of applicants per vacancy was under two.

Why are there people who are unemployed when there are

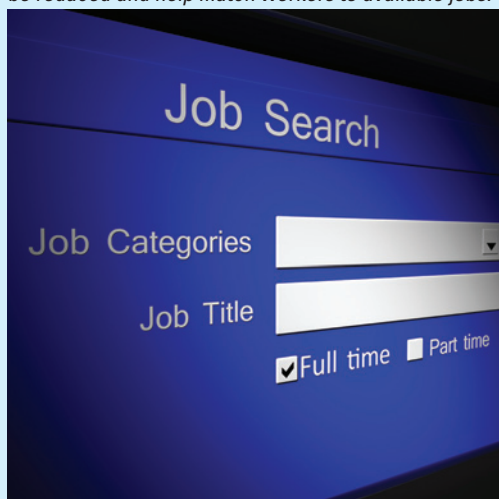
sometimes large numbers of vacant positions waiting to be filled? This was a question asked by three economists who were awarded the Nobel Prize for Economics in 2010. Peter A. Diamond, Dale T. Mortenson and Christopher A. Pissarides developed a theory known as 'search and matching theories'.

Under normal circumstances, markets are assumed to be in equilibrium when supply equals demand. When something happens to disturb that equilibrium, market forces work to restore equilibrium. The work of these three economists has shown how labour markets can operate differently and that unemployment can remain at high levels even though

there are plenty of vacancies. Government policy and labour market regulation may have important consequences for labour markets and can contribute to the disequilibrium rather than helping to clear the market.

Some of the reasons include the time it takes for those who are unemployed to find work – there are search costs involved. For employers there are transaction costs to hiring labour and whilst there may be employers looking for labour and labour looking for work, both parties need to be put in touch with each other and agree to a transaction. Even if the two parties do come together a transaction may not be guaranteed (i.e. the supplier of labour gets a job) because there may be incentives for both sides to see if they can get a better outcome for themselves. In addition, policies regulating the ease with which employers can hire or fire workers and the benefits or unemployment insurance that exists in a country can also influence when labour

In reality the labour market is not perfect and is subject to a number of frictions. The development of job search sites on the Internet may be one way in which search costs for employers and employees can be reduced and help match workers to available jobs.



becomes unemployed and how long they stay unemployed. If for example, unemployment benefit is relatively generous then it may be that workers take more time to find work because they have some degree of support to live on whilst they find 'the right job'. It is entirely possible, therefore, that vacancies can remain unfilled even when there are high levels of unemployment and that the more generous unemployment benefits are the higher unemployment might be.

The search process, therefore, involves both time and resources for both employers and the unemployed looking for work and these create frictions in the market which mean that the wants of both employers and the unemployed remain unmet. Both could agree on a price if they were able to meet and thus fulfil one of the basic requirements for a functioning market, but the search frictions mean that the labour market will fail to clear at particular points in time. Search theory provides a departure point from the assumptions of perfect competition in the labour market.

The theory further suggests that when the economy is subject to a shock, such as the financial crisis or recession, it will take much longer for unemployment to return to pre-shock levels even though firms may be looking to hire workers once again.

The research carried out by the laureates is helping economists and policymakers gain a better understanding of policies to help reduce high levels of

unemployment. They have brought a different perspective to the debate on how best to deal with the frictional aspects of structural unemployment, which seems to be a problem for a number of countries following the financial crisis where jobs have been lost in one sector but workers have found difficulty moving to other sectors of the economy where vacancies exist.

Questions

- 1 **What do you think is the main cause of the differences in regional unemployment implied by the different applicant to vacancy rates quoted for the UK: frictional or structural? Explain.**
- 2 **Why is it important to identify the type of unemployment and the nature of the unemployed when looking at levels of unemployment?**
- 3 **The number of unemployed is much higher in the UK than the other countries quoted in the article and the proportion of vacancies to the unemployed also much higher. What questions would you ask if analysing these data to discover why this is the case?**
- 4 **Why might search costs render some of the assumptions of perfect competition redundant and thus result in the labour market not clearing?**
- 5 **To what extent can governments reduce search costs or does government intervention simply exacerbate frictions in the market?**

SUMMARY

- The unemployment rate is the percentage of those who would like to work who do not have jobs. The government calculates this statistic monthly based on a survey of thousands of households.
- The unemployment rate is an imperfect measure of joblessness. Some people who call themselves unemployed may actually not want to work, and some people who would like to work have left the labour force after an unsuccessful search.
- In many advanced economies, most people who become unemployed find work within a short period of time. Nevertheless, most unemployment observed at any given time is attributable to the few people who are unemployed for long periods of time.
- One reason for unemployment is the time it takes for workers to search for jobs that best suit their tastes and skills. Unemployment insurance is a government policy that, while protecting workers' incomes, increases the amount of frictional unemployment.
- A second reason why an economy may always have some unemployment is if there is a minimum wage that exceeds the wage that would balance supply and demand for the workers who are eligible for the minimum wage. By raising the wage of unskilled and inexperienced workers above the equilibrium level, minimum wage laws raise the quantity of labour supplied and reduce the quantity demanded. The resulting surplus of labour represents unemployment.
- A third reason for unemployment is the market power of unions. When unions push the wages in unionized industries above the equilibrium level, they create a surplus of labour.
- A fourth reason for unemployment is suggested by the theory of efficiency wages. According to this theory, firms find it profitable to pay wages above the equilibrium level. High wages can improve worker health, lower worker turnover, increase worker effort and raise worker quality.
- The costs of unemployment to individuals include lower incomes, loss of self-esteem, de-skilling, an increase in the possibility of mental health problems, family breakdown and crime.
- The costs of unemployment to society as a whole includes the opportunity cost of lost output which reduces potential growth, the impact on government income and expenditure, the social costs of dealing with unemployment and its effects, the underuse of resources and the reverse multiplier effect.

QUESTIONS FOR REVIEW

- 1 Why might a person who does not have a job not be unemployed?
- 2 How do national statistical offices compute the labour force, the unemployment rate and the labour force participation rate?
- 3 Is unemployment typically short term or long term? Explain.
- 4 Why is frictional unemployment inevitable? How might the government reduce the amount of frictional unemployment?
- 5 Are minimum wage laws a better explanation for structural unemployment among teenagers or among university graduates? Why?
- 6 How do unions affect the natural rate of unemployment?
- 7 What is the hysteresis effect?
- 8 What claims do advocates of unions make to argue that unions are good for the economy?
- 9 Explain four ways in which a firm might increase its profits by raising the wages it pays.
- 10 Outline three costs of unemployment to (a) an individual and (b) to society.

PROBLEMS AND APPLICATIONS

- 1 Assume that in a country, of all adult people, 29,500,000 were employed, 1,120,000 were unemployed and 11,000,000 were not in the labour force. How big was the labour force? What was the labour force participation rate? What was the unemployment rate?
- 2 Go to the website of the UK Office for National Statistics (<http://www.statistics.gov.uk>) or Eurostat (<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>) or the statistics office for the country in which you are studying. What is the national unemployment rate right now? Find the unemployment rate for the demographic

- group that best fits a description of you (for example, based on age, sex and ethnic group). Is it higher or lower than the national average? Why do you think this is so?
- 3 According to a Labour Force Survey in a country, total employment increased by around 1.35 million workers between 2010 and 2016, but the number of unemployed workers increased by 350,000.
 - a. How are these numbers consistent with each other?
 - b. In some cases, there can be an increase in the number of people employed but the number of unemployed people declines by a smaller amount. Why might one expect a reduction in the number of people counted as unemployed to be smaller than the increase in the number of people employed?
 - 4 Are the following workers more likely to experience short-term or long-term unemployment? Explain.
 - a. A construction worker laid off because of bad weather.
 - b. A manufacturing worker who loses her job at a plant in an isolated area.
 - c. A bus industry worker laid off because of competition from the railway.
 - d. A short-order cook who loses his job when a new restaurant opens across the street.
 - e. An expert welder with little formal education who loses her job when the company installs automatic welding machinery.
 - 5 Using a diagram of the labour market, show the effect of an increase in the minimum wage on the wage paid to workers, the number of workers supplied, the number of workers demanded and the amount of unemployment.
 - 6 Do you think that firms in small towns or in cities have more market power in hiring? Do you think that firms generally have more market power in hiring today than 50 years ago, or less? How do you think this change over time has affected the role of unions in the economy? Explain.
 - 7 Consider an economy with two labour markets, neither of which is unionized. Now suppose a union is established in one market.
 - a. Show the effect of the union on the market in which it is formed. In what sense is the quantity of labour employed in this market an inefficient quantity?
 - b. Show the effect of the union on the non-unionized market. What happens to the equilibrium wage in this market?
 - 8 Some workers in the economy are paid a flat salary and some are paid by commission. Which compensation scheme would require more monitoring by supervisors? In which case do firms have an incentive to pay more than the equilibrium level (as in the worker effort variant of efficiency wage theory)? What factors do you think determine the type of compensation firms choose?
 - 9 Suppose that the government passes a law requiring employers to provide employees some benefit (such as a guaranteed pension) that raises the cost of an employee by €4 per hour.
 - a. What effect does this new law have on the demand for labour? (In answering this and the following questions, be quantitative when you can.)
 - b. If employees place a value on this benefit exactly equal to its cost, what effect does the new law have on the supply of labour?
 - c. If the wage is free to balance supply and demand, how does this law affect the wage and the level of employment? Are employers better or worse off? Are employees better or worse off?
 - d. If a minimum wage law prevents the wage from balancing supply and demand, how does the new law affect the wage, the level of employment and the level of unemployment? Are employers better or worse off? Are employees better or worse off?
 - e. Now suppose that workers do not value the benefit arising from the new law at all. How does this alternative assumption change your answers to parts (b), (c) and (d) above?
 - 10 Statistics from the International Labour Organization in 2011, suggest that there are 75 million young people (those aged between 16 and 23) unemployed worldwide. Young people are three times more likely to be unemployed than those in other age groups and each month out of work prior to age 23 increases the likelihood that they will experience longer periods of unemployment later in life.
 - a. Why are such statistics of importance to policymakers?
 - b. What specific costs might be associated with youth unemployment to young people and to society?
 - c. Should policymakers devote more resources to dealing with youth unemployment compared to other groups in society? Justify your response.

PART 12

INTEREST RATES, MONEY AND PRICES IN THE LONG RUN

24

SAVING, INVESTMENT AND THE FINANCIAL SYSTEM

Imagine that you have just graduated from university (with a degree in economics, of course) and you decide to start your own business – an economic forecasting firm. Before you make any money selling your forecasts, you have to incur substantial costs to set up your business. You have to buy computers with which to make your forecasts, as well as desks, chairs and filing cabinets to furnish your new office. Each of these items is a type of capital that your firm will use to produce and sell its services.

How do you obtain the funds to invest in these capital goods? Perhaps you are able to pay for them out of your past savings. More likely, however, like most entrepreneurs, you do not have enough money of your own to finance the start of your business. As a result, you have to get the money you need from other sources.

There are various ways for you to finance these capital investments. You could borrow the money, perhaps from a bank or from a friend or relative. In this case, you would promise not only to return the money at a later date but also to pay interest for the use of the money. Alternatively, you could convince someone to provide the money you need for your business in exchange for a share of your future profits, whatever they might happen to be. In either case, your investment in computers and office equipment is being financed by someone else's saving.

The **financial system** consists of those institutions in the economy that help to match one person's saving with another person's investment. As we discussed in the previous chapter, saving and investment are key ingredients to long-run economic growth: when a country saves a large portion of its GDP, more resources are available for investment in capital, and higher capital raises a country's productivity and living standard. The previous chapter, however, did not explain how the economy coordinates saving and investment. At any time, some people want to save some of their income for the future, and others want to borrow in order to finance investments in new and growing businesses. What brings these two groups of people together? (Not physically of course!) What ensures that the supply of funds from those who want to save balances the demand for funds from those who want to invest?

financial system the group of institutions in the economy that help to match one person's saving with another person's investment

This chapter examines how the financial system works. First, we discuss the large variety of institutions that make up the financial system. Second, we discuss the relationship between the financial system and some key macroeconomic variables – notably saving and investment. Third, we develop a model of the supply and demand for funds in financial markets. In the model, the interest rate is the price that adjusts to balance supply and demand. The model shows how various government policies affect the interest rate and, thereby, society's allocation of scarce resources.

FINANCIAL INSTITUTIONS IN THE ECONOMY

At the broadest level, the financial system moves the economy's scarce resources from savers (people who spend less than they earn) to borrowers (people who spend more than they earn). Savers save for various reasons – to help put a child through university in several years or to retire comfortably in several decades. Similarly, borrowers borrow for various reasons – to buy a house in which to live or to start a business with which to make a living. Savers supply their money to the financial system with the expectation that they will get it back with interest at a later date. The expectation that the reward for saving will be the receipt of interest is not universal – Islamic finance has a very different approach to finance but in this chapter we will be focusing on what might be called 'traditional' financial institutions. Borrowers demand money from the financial system with the knowledge that they will be required to pay it back with interest at a later date.

The financial system is made up of various financial institutions that help coordinate savers and borrowers. Financial institutions can be grouped into two categories – financial markets and financial intermediaries. We consider each category in turn.

Financial Markets

Financial markets are the institutions through which a person who wants to save can directly supply funds to a person who wants to borrow. Two of the most important financial markets in advanced economies are the bond market and the stock market.

financial markets financial institutions through which savers can directly provide funds to borrowers

The Bond Market When BP, the oil company, wants to borrow to finance a major new oil exploration project, it can borrow directly from the public. It does this by selling bonds. A **bond** is a certificate of indebtedness that specifies the obligations of the borrower to the holder of the bond. Put simply, a bond is an IOU. It identifies the time at which the loan will be repaid, called the *date of maturity*, and the rate of interest that will be paid periodically (called the *coupon*) until the loan matures. The buyer of a bond gives his or her money to BP in exchange for this promise of interest and eventual repayment of the amount borrowed (called the *principal*). The buyer can hold the bond until maturity or can sell the bond at an earlier date to someone else.

bond a certificate of indebtedness

There are literally millions of bonds traded in advanced economies. When large corporations or the national government, or even local governments need to borrow in order to finance the purchase of a new factory, a new jet fighter or a new school, they often do so by issuing bonds. If you look at the *Financial Times* or the business section of any national newspaper, you will find a listing of the prices and interest rates on some of the most important bond issues. Although these bonds differ in many ways, two characteristics of bonds are most important.

The first characteristic is a bond's *term* – the length of time until the bond matures. Some bonds have short terms, such as a few months, while others have terms as long as 30 years. (The British government has even issued a bond that never matures, called a *perpetuity*. This bond pays interest forever, but the principal is never repaid.) The interest rate on a bond depends, in part, on its term. Long-term bonds are riskier than short-term bonds because holders of long-term bonds have to wait longer for repayment of the principal. If a holder of a long-term bond needs his money earlier than the distant date of maturity, he has no choice but to sell the bond to someone else, perhaps at a reduced price. To compensate for this risk, long-term bonds usually (but not always) pay higher interest rates than short-term bonds.

The second important characteristic of a bond is its *credit risk* – the probability that the borrower will fail to pay some of the interest or principal. Such a failure to pay is called a *default*. Borrowers can (and sometimes do) default on their loans by declaring bankruptcy. When bond buyers perceive that the probability of default is high, they demand a higher interest rate to compensate them for this risk. When national governments want to borrow money to finance public spending, they issue bonds. You will hear this referred to as *sovereign debt*. Some government bonds are considered a safe credit risk, such as those from Germany, for example, and tend to pay low interest rates. UK government bonds have come to be referred to as *gilt-edged* bonds, or more simply as *gilts*, reflecting that, in terms of credit risk, they are 'as good as gold' (early bond certificates had a gold edge – hence the term 'gilt edged'). In contrast, financially shaky corporations raise money by issuing *junk bonds*, which pay very high interest rates; in recent years some countries' debt has been graded as 'junk'. Buyers of bonds can judge credit risk by checking with various private agencies, such as Standard & Poor's, which rate the credit risk of different bonds. Sometimes, these bonds are referred to euphemistically but less graphically as *below investment grade bonds*.

One important point to note is the relationship between a bond's price and its yield. Assume that a corporation issues a €1,000 bond over a ten-year period with a coupon of 3.5 per cent. For the duration of the ten years, the corporation will pay the bond holder €35 a year in interest and when the bond matures the corporation will pay back the bond holder the €1,000 principal. At any time during the ten-year period the bond holder can sell the bond on the bond market. The price she gets will depend on the supply and demand of those bonds on the market. The yield of the bond (in simple terms) is given by the coupon/price $\times 100$. Price is quoted as a percentage of the principal. Assume the bond holder needs to get access to cash (liquidity) quickly – they decide to sell their bond. The price of the bond on the market is €995.

The yield on this bond, therefore, is $35/995 \times 100 = 3.52$ per cent. If the seller was able to sell the bond for €1,050, the yield would be $35/1050 \times 100 = 3.33$ per cent. This is an important point about the price and yield of a bond – there is an inverse relationship between price and yield. As bond prices rise, the yield falls and vice versa. Why might bond prices rise and fall on the markets? The demand and supply of bonds (the number of people wanting to buy bonds and the number of people wanting to sell bonds) is made up of existing bonds and the availability of new bonds. Bond prices (and, therefore, yields) are affected by existing bonds in the market, the issue of new bonds, the likelihood of the bond issuer defaulting and the interest rates on other securities. The issue of a new bond is also affected by these factors. If current interest rates are high, new issues have to have a coupon which will compete and vice versa.

The Stock Market Another way for BP to raise funds for its oil exploration project is to sell stock in the company. **Stock** represents ownership in a firm and is, therefore, a claim to the future profits that the firm makes. For example, if BP sells a total of 1,000,000 shares of stock, then each share represents ownership of 1/1,000,000 of the business. A stock is also commonly referred to as a *share* or as an *equity*. In this book, we'll use the terms 'share' and 'stock' (and 'stockholder' and 'shareholder') more or less interchangeably.

stock (or share or equity) a claim to partial ownership in a firm

The sale of stock to raise money is called *equity finance*, whereas the sale of bonds is called *debt finance*. Although corporations use both equity and debt finance to raise money for new investments, stocks and bonds are very different. The owner of BP shares is a part owner of BP; the owner of a BP bond is a creditor of the corporation. If BP is very profitable, the shareholders enjoy the benefits of these profits, whereas the bondholders get only the interest on their bonds. And if BP runs into financial difficulty, the

bondholders are paid what they are due before shareholders receive anything at all. Compared to bonds, stocks offer the holder both higher risk and potentially higher return.

As we noted just now, stocks are also called shares or equities. In the UK, bonds are also, confusingly, referred to as 'stock'. This term for government bonds has been in use in England since the late 17th century and is well established. In order to avoid confusion, however, the term is often qualified as *government stock* or *gilt-edged stock*. In general, though, despite the confusing use of language, the term *stock* refers to ownership of a firm.

After a corporation issues stock by selling shares to the public, these shares trade among stockholders on organized stock exchanges. In these transactions, the corporation itself receives no money when its stock changes hands. Most of the world's countries have their own stock exchanges on which the shares of national companies trade.

The prices at which shares trade on stock exchanges are determined by the supply and demand for the stock in these companies. Because stock represents ownership in a corporation, the demand for a stock (and thus its price) reflects people's perception of the corporation's future profitability. When people become optimistic about a company's future, they raise their demand for its stock and thereby bid up the price of a share of stock. Conversely, when people come to expect a company to have little profit or even losses, the price of a share falls.

Various stock indices are available to monitor the overall level of stock prices for any particular stock market. A *stock index* is computed as an average of a group of share prices. The Dow Jones Industrial Average has been computed regularly for the New York Stock Exchange since 1896. It is now based on the prices of the shares of 30 major US companies. The Financial Times Stock Exchange (FTSE) 100 Index, is based on the top 100 companies (according to the total value of their shares) listed on the London Stock Exchange (LSE), while the FTSE All-Share Index is based on all companies listed on the LSE. Indices of prices on the Frankfurt stock market, based on 30 and 100 companies respectively, are the DAX 30 and DAX 100. The NIKKEI 225 (or just plain NIKKEI Index) is based on the largest 225 companies, in terms of market value of shares, traded on the Tokyo Stock Exchange.

Because share prices reflect expected profitability, stock indices are watched closely as possible indicators of future economic conditions.

Financial Intermediaries

Financial intermediaries are financial institutions through which savers can indirectly provide funds to borrowers. The term *intermediary* reflects the role of these institutions in standing between savers and borrowers. Here we consider two of the most important financial intermediaries – banks and investment funds.

financial intermediaries financial institutions through which savers can indirectly provide funds to borrowers

Banks If the owner of a small business wants to finance an expansion of his business, he probably takes a strategy quite different from BP. Unlike BP, a small businessman would find it difficult to raise funds in the bond and stock markets. Most buyers of stocks and bonds prefer to buy those issued by larger, more familiar companies. The small businessman, therefore, most likely finances his business expansion with a loan from a bank.

Banks are the financial intermediaries with which people are most familiar. A primary function of banks is to take in deposits from people who want to save and use these deposits to make loans to people who want to borrow. Banks pay depositors interest on their deposits and charge borrowers slightly higher interest on their loans. The difference between these rates of interest covers the banks' costs and returns some profit to the owners of the banks.

Besides being financial intermediaries, banks play a second important role in the economy: they facilitate purchases of goods and services by allowing people to write cheques against their deposits (although cheques are being phased out in many countries as digital payment systems become more common), or to use debit cards to transfer money electronically from their account to the account of the person or corporation they are buying something from. In other words, banks help create a special asset that people

can use as a *medium of exchange*. A medium of exchange is an item that people can easily use to engage in transactions. A bank's role in providing a medium of exchange distinguishes it from many other financial institutions. Stocks and bonds, like bank deposits, are a possible *store of value* for the wealth that people have accumulated in past saving, but access to this wealth is not as easy, cheap and immediate as just writing a cheque or swiping a debit card. For now, we ignore this second role of banks, but we will return to it when we discuss the monetary system later in the book.

Investment Funds A financial intermediary of increasing importance is the investment fund. An **investment fund** is an institution that sells shares to the public and uses the proceeds to buy a selection, or *portfolio*, of various types of shares, bonds, or both shares and bonds. The shareholder of the investment fund accepts all the risk and return associated with the portfolio. If the value of the portfolio rises, the shareholder benefits; if the value of the portfolio falls, the shareholder suffers the loss.

investment fund an institution that sells shares to the public and uses the proceeds to buy a portfolio of stocks and bonds

The primary advantage of investment funds is that they allow people with small amounts of money to diversify. Buyers of shares and bonds are well advised to heed the adage, 'Don't put all your eggs in one basket.' Because the value of any single stock or bond is tied to the fortunes of one company, holding a single kind of stock or bond is very risky. By contrast, people who hold a diverse portfolio of shares and bonds face less risk because they have only a small stake in each company. Investment funds make this diversification easy. With only a few hundred euros, a person can buy shares in an investment fund and, indirectly, become the part owner or creditor of hundreds of major companies. For this service, the company operating the investment fund charges shareholders a fee, usually between 0.5 and 2.0 per cent of assets each year. Closely related to investment funds are unit trusts, the difference being that when people put money into a unit trust, more 'units' or shares are issued, whereas the only way to buy into an investment fund is to buy existing shares in the fund. For this reason, unit trusts are sometimes referred to as 'open-ended'.

A second advantage claimed by investment fund companies is that investment funds give ordinary people access to the skills of professional money managers. The managers of most investment funds pay close attention to the developments and prospects of the companies in which they buy stock. These managers buy the stock of those companies that they view as having a profitable future and sell the stock of companies with less promising prospects. This professional management, it is argued, should increase the return that investors in investment funds earn on their savings.

Financial economists, however, are often sceptical of this second argument. With thousands of money managers paying close attention to each company's prospects, the price of a company's stock could be seen as being a good reflection of the company's true value. As a result, it is hard to 'beat the market' by buying good stocks and selling bad ones. In fact, investment funds called *index funds* or *tracker trusts* which buy all (or the large majority of) the stocks in a given stock index, perform somewhat better on average than investment funds that take advantage of active management by professional money managers. The explanation for the superior performance of tracker trusts is that they keep costs low by buying and selling very rarely and by not having to pay the salaries of the professional money managers.

SELF TEST What is stock? What is a bond? How are they different? How are they similar?

Credit Default Swaps (CDS)

The financial crisis of 2007–2009 highlighted the role of the range of financial instruments traded in the financial system. These instruments play a role in the financial system but have become increasingly complex; indeed, there were some senior bankers who were accused of not fully understanding what they were trading! The first of these instruments are credit default swaps or CDS. A **credit default swap** is a means by which bondholders can insure themselves against the risk of default.

credit default swap (CDS) a means by which a bondholder can insure against the risk of default

Whenever a bond is sold there is an associated risk attached to it. In the case of bonds backed by a pool of mortgage debt, the risk is that the mortgage payer defaults on the payment in some way. CDS are a means of insuring against the risk involved. To see how this works let us take an example.

Assume a bank has bought a bond in an asset-backed security (in reality of course it might be many bonds) worth €5 million. The bond has a principal, therefore, of €5 million. The coupon payment (the interest on the bond) is 10 per cent. The bond is backed by the stream of cash flows being paid by mortgage holders – the underlying asset to the bond. The bondholder knows that there is a risk that some mortgage holders might default and as a result the bond return might also be in default – i.e. it does not pay the coupon or possibly the principal. If mortgage holders all meet their obligations and pay the money back (the principal) the bank will have earned interest for as long as it held the bond. If the bond was a ten-year bond with a coupon of 10 per cent then the bondholder would have earned a 10 per cent return on the bond for ten years (€500,000 per year for ten years).

The bondholder can go to an insurance company (which could also be another bank or a hedge fund) and take out a policy on the risk. The bank seeking to insure the risk is referred to as the *protection buyer* and the insurance company or other financial institution, the *protection seller*. The policy will agree to restore the bondholder to their original position should the bond default. The bondholder pays the insurance company premiums over the life of the policy, let us assume this is €500,000 a year over ten years, and in return the insurance company would agree to pay the bondholder €5 million if the bond defaults in some way during the period of the agreement.

What the bondholder has done is to swap the risk in the bond with the insurance company. In the event of the bond defaulting then the protection seller pays out to the protection buyer. These CDS could also be traded and in many cases the protection buyer might take out a number of such hedges against its risk.

The price of a CDS is quoted in terms of basis points where a basis point is $\frac{1}{100}$ th of 1 per cent. The price quoted reflects the cost of protecting debt for a period of time. In May 2010, for example, the price of a CDS for BP was quoted at 255 bps suggesting that the cost of protecting €10 million of BP debt for five years was €255,000 a year.

If the bond was subject to default this would trigger a payment called a *credit event*. If a credit event occurs then there are two ways in which the contract could be closed. One was through a physical settlement. In this case the protection buyer would deliver the bond to the seller who would give the buyer the value of the bond, called par. The CDS market expanded rapidly throughout the 'noughties' leading to multiple trading. The holder of the CDS might not actually own the bond in question and so this method of settlement clearly does not work. The growth in the market meant that the number of contracts far outweighed the number of bonds on which they were based. To settle the claims of contract holders through physical means would not be possible. As a consequence, protection sellers undertook to settle the contract claim through cash settlement where the positions were cleared by the transfer of cash.

The protection seller would also have to put up some form of collateral to cover its exposure to the possible default as part of the CDS agreement. The size of the exposure in the example above would be €5 million. This would be part of the contract between the insurance seller and the buyer. The amount of the collateral required would depend on the market value of the bond and the credit rating of the protection seller. The market value of the bond can change because the bondholder can sell their bond.

If a bond associated with a mortgage pool is deemed a higher risk then it may become more difficult to sell it and its price would fall. At a lower price the risk involved for the protection buyer is higher and they can exercise a call (known as a collateral call) on the protection seller to increase the collateral supplied. If, for example, the price of the bond in the example above fell to 40 cents in the euro, the insurance company would have to find $0.60 \times €500,000$ in additional collateral (€3 million). This process is known as collateral calls. Equally, if the rating of the protection seller is downgraded then the risk to the protection buyer also rises – the seller may be not be able to meet its obligations. Once again, the protection buyer

can call the seller to ask for more collateral to be provided. The insurance company would then have to find more funds to top up the collateral it has to provide to reflect the lower value of the bond.

If the seller was unable to meet its obligations at least the buyer would have some protection in the form of the collateral supplied by the seller. If the bond runs its course and matures without any credit event then the collateral is returned to the seller (who also, remember, receives premiums from the buyer).

CDS represent sound business principles when the risk of default is very low, which was the case when they were first developed in the late 1990s and when the bonds being insured were corporate bonds with a low risk attached to them. The expansion of CDS presented a different challenge to the financial markets when the housing market collapse occurred from around 2007.

CASE STUDY Ireland and the Financial Crisis

One of the first countries to be affected by the financial crisis of 2007/08 was Ireland. In the 1990s and early part of the 2000s, Ireland's economy grew at rates of between 4.4 per cent and 6.5 per cent throughout the years from 2000 to 2007, fuelled in part by boom in property prices (the construction industry accounted for around a quarter of Ireland's GDP in 2007), advantageous tax rates and the benefits it was gaining from being a member of the European Union. When the financial crisis struck, the Irish economy was hit hard, shrinking by 3 per cent in 2008 and 7 per cent in 2009. The Irish banking system was particularly badly hit and the Irish government had to ask the EU for a bailout to help support the banking system.

Support from the EU and the International Monetary Fund (IMF) has amounted to some €85 billion but securing this support has come at a price – the cuts in government spending and austerity measures that the government has put in place have increased unemployment to a peak of 15 per cent and seen many people suffer as property prices collapsed. As the austerity plans have taken effect the prospects for the Irish economy were starting to look a little brighter in the latter half of 2012 and into 2013. One example of this brighter future was the re-entry of the Irish government into the sovereign bond market. In July 2012, the Irish government managed to raise €500 million in an auction that was oversubscribed by 2.8 times and represented the first time the Irish government has ventured into the capital markets to borrow money since 2010. At the auction, the government borrowed through the issue of three-month treasury bills with a yield of 1.8 per cent.

In early 2013, Ireland managed to raise €2.5 billion through a bond sale, which was again oversubscribed by more than two-and-a-half-times, at a yield of around 3.35 per cent. Irish bond yields had been gradually falling since 2011 when a peak of 14 per cent was reached suggesting that investors have more confidence in the Irish government and the economy. The Irish government will see the support from the capital markets as encouraging because it was planning to raise €10 billion in bond issues in 2013 to help fund government spending and emerge from the funding support it has received from the EU and IMF. In the early part of 2013, the Irish government was on course to meet the projected targets for reducing its budget deficits and this has helped boost confidence in the markets along with the gradual improvement in the economy as a whole.



Austerity policies have affected a lot of people in Ireland but the economy showed signs of improvement in 2013. The cost of turning round the economy has been very high for many individuals and businesses, however.

Collateralized Debt Obligations (CDOs)

CDOs are pools of asset-backed securities which are dependent on the value of the asset that backs them up and the stream of income that flows from these assets. Essentially, CDOs work in the following way. In setting up a CDO, a manager encourages investors to buy bonds, the funds of which are used to buy pools of debt – mortgage debt. This debt is split and rated according to its risk into tranches; low-risk tranches attract low interest rates whilst the riskier tranches attract higher interest rates. Under ‘*normal*’ circumstances (the relevance of the emphasis will become clear later) the payments by mortgage holders provide sufficient income each month to pay the interest to each of the tranche holders. There is a risk, of course, that some of the mortgage holders in the initial debt could default on payment but historical data enable investors to have some idea of what this risk will be. In the event of default, some of the riskier tranches may not get paid – that is the risk they take and why they get a higher interest rate.

The development of new approaches to risk management enabled these structures to be extended further into second and third ‘waves’ of securitized debt. Asset managers could buy particular tranches of debt (backed, remember, by mortgages ultimately) and mix them with other types of debt and sell them on to other investors. Investors in these higher risk tranches are assured by the ratings on the investment. Calculation of default correlations during the latter part of the 1990s and into the first years of the 2000s was relatively low but when times were not ‘*normal*’ the correlations could start to become far more unstable. Problems for CDOs began when holders of sub-prime mortgages began to default on payments. The **sub-prime market** offered mortgage opportunities to those not traditionally seen as being part of the financial markets because of their high credit risk and was part of the way in which banks and other lenders sought to increase their lending. Individuals seen as being a relatively low credit risk for mortgage lending were known as the prime market; the term is said to have derived from analogy with the best cuts of meat from an animal. It followed that there was a sub-prime group for whom access to mortgages was altogether more difficult.

sub-prime market individuals not traditionally seen as being part of the financial markets because of their high credit risk

Some of these people had credit histories that were very poor, some did not have jobs, but in an atmosphere of risk seeking and changed priorities this group provided lenders with a market opportunity because they were willing to pay high rates of interest on their mortgages. At first these defaults might have been ‘pin pricks’ in the structure but as the number increased the payments to the first tranches began to reduce. Those at the top of the ratings list may have still got paid but the riskier lower tranches began to see their payments dry up. In turn, the asset-backed securities sold in the second and subsequent ‘waves’ began to see their payments cease and as the sub-prime problem became worse it became clear that it was likely that these subsequent ‘waves’ would not be paid back and were thus worthless. Holders of this debt (including banks, governments and local governments around the world) found that they had to write-off large assets. The term ‘toxic assets’ became familiar to many who may never have thought they were affected by high finance. These were mortgage-backed securities and other debt (such as bonds) that are not able to be repaid in many cases because the value of the assets against which they are secured have fallen significantly.

Summing Up

An advanced economy contains a large variety of financial institutions. In addition to the bond market, the stock market, banks and investment funds, there are also pension funds, insurance companies and – notably in the UK and USA – even the local pawnbroker’s shop: places where people who might be deemed bad credit risks can borrow money and leave some valuable item such as a watch or a piece of jewellery as security in case the loan is not repaid. Clearly, these institutions differ in many ways and the products they sell have become more complex in lots of cases. When analysing the macroeconomic role of the financial system, however, it is more important to keep in mind the similarity of these institutions than the differences. These financial institutions all serve the same goal – directing the resources of savers into the hands of borrowers.

FYI



Dark Pools

Despite having a rather sinister sounding name, dark pools are just an electronic network which puts buyers of shares in touch with sellers. However, the nature of dark pools has raised some concerns and is currently being looked at by the Securities and Exchange Commission (SEC) in the United States (which may account for why the volume of shares traded through these systems fell in 2012). Dark pools are not widespread in the UK and Europe at the time of writing, but are likely to become more significant in the future.

How do dark pools work? Under normal circumstances, trades in shares are carried out via a number of exchanges such as the London and New York stock exchanges. Traders on those exchanges can see who is doing the buying and who is doing the selling. Any transaction that goes through, therefore, can affect the market and traders will be able to amend their positions accordingly. Dark pools allow trades to be carried out anonymously so the trade can go through but no one knows – crucially at the time of the trade – who is doing the buying and selling. Dark pools tend to be used for trades where significant volumes of shares are transacted. The following is an outline of how it works.

Assume an institutional investor, such as a pension fund or insurance company, has \$1 billion worth of shares in ExxonMobil that it wants to sell. It knows that if it attempts to trade this amount it is likely to affect the market. Typically, this sort of volume would have to be broken up and sold in chunks. If the price of ExxonMobil shares was currently \$500 each, then as the sale progresses, the insurance company would end up with lower prices for the latter chunks of shares than the first lots. The average price it gains from the sale would, therefore, be less than \$500 each.

The price would fall because the supply of ExxonMobil shares would be rising, traders would see the trades going through and adjust their positions accordingly. If the insurance company carries out the trade via a dark pool it places the shares into an electronic exchange. In the pool there are buyers and sellers all looking for trades, but anonymously. The use of algorithms to improve the efficiency of the trades in dark pools is common. The algorithm is programmed to search for buyers who want to purchase \$1 billion of ExxonMobil shares, for example, and matches the two together. At this point in the trade no-one knows who the traders are. Once the trade is completed then it is announced.

The announcement may, or may not, have an effect on the market. The advantage for the pension fund or insurance company is that it completes the trade for the volume of shares it wants to deal and the use of dark pools prevents the market reacting to the sale at the time it is happening and thus affecting the price the seller gets.

There has been some criticism of dark pools mainly because of two key issues. It is feared that dark pools may take too much trade away from traditional exchanges. As they develop, more trading will be conducted in secret and in some cases, even though the trades are announced, the identity of the trader is not revealed. Dark pools may discriminate against individual investors who can be affected by the price changes that result from the activity which they could not foresee because of the very nature of the anonymous transactions.

Despite the criticism, proponents of dark pools point to the efficiency with which they operate and the fact that they help liquidity in the markets. The bank Goldman Sachs has been criticized for its use of dark pools but has defended the practice and suggested that enhanced transparency post-trade is an obvious way to offset concern about secrecy.

SAVING AND INVESTMENT IN THE NATIONAL INCOME ACCOUNTS

Events that occur within the financial system are central to understanding developments in the overall economy. As we have just seen, the institutions that make up this system – the bond market, the stock market, banks and investment funds – have the role of coordinating the economy's saving and investment. And as we saw in the previous chapter, saving and investment are important determinants of long-run growth in GDP and living standards. As a result, macroeconomists need to understand how financial markets work and how various events and policies affect them.

As a starting point for an analysis of financial markets, we discuss in this section the key macroeconomic variables that measure activity in these markets. Our emphasis here is not on behaviour but on accounting. *Accounting* refers to how various numbers are defined and added up. A personal accountant might help an individual add up his income and expenses. A national income accountant does the same thing for the economy as a whole. The national income accounts include, in particular, GDP and the many related statistics.

The rules of national income accounting include several important identities. Recall that an *identity* is an equation that must be true because of the way the variables in the equation are defined. Identities are useful to keep in mind, for they clarify how different variables are related to one another. Here we consider some accounting identities that shed light on the macroeconomic role of financial markets.

Some Important Identities

Recall that GDP is both total income in an economy and the total expenditure on the economy's output of goods and services. GDP (denoted as Y) is divided into four components of expenditure: consumption (C), investment (I), government purchases (G) and net exports (NX). We write:

$$Y = C + I + G + NX$$

This equation is an identity because every euro of expenditure that shows up on the left-hand side also shows up in one of the four components on the right-hand side. Because of the way each of the variables is defined and measured, this equation must always hold. Sometimes we make this clear by using an identity sign, with three bars, instead of the usual equals sign with two bars:

$$Y \equiv C + I + G + NX$$

In general, though, we can use the usual equality sign.

In this chapter, we simplify our analysis by assuming that the economy we are examining is closed. A *closed economy* is one that does not interact with other economies. In particular, a closed economy does not engage in international trade in goods and services, nor does it engage in international borrowing and lending. Of course, actual economies are *open economies* – that is, they interact with other economies around the world. Nevertheless, assuming a closed economy is a useful simplification with which we can learn some lessons that apply to all economies. Moreover, this assumption applies perfectly to the world economy (since interplanetary trade is not yet common).

Because a closed economy does not engage in international trade, imports and exports are exactly zero. Therefore, net exports (NX) are also zero. In this case, we can write:

$$Y = C + I + G$$

This equation states that GDP is the sum of consumption, investment and government purchases. Each unit of output sold in a closed economy is consumed, invested or bought by the government.

To see what this identity can tell us about financial markets, subtract C and G from both sides of this equation. We obtain:

$$\begin{aligned} Y - C - G &= (C - C) + I + (G - G) \\ Y - C - G &= I \end{aligned}$$

The left-hand side of this equation ($Y - C - G$) is the total income in the economy that remains after paying for consumption and government purchases: this amount is called **national saving**, or just saving, and is denoted S . Recall from the more complex circular flow of income diagram in Chapter 20, that there are withdrawals and injections from the circular flow. Saving is one of the withdrawals from the circular flow but reappears in the economy in the form of investment when savings are channelled by financial institutions.

national saving (saving) the total income in the economy that remains after paying for consumption and government purchases

Substituting S for $Y - C - G$, we can write the last equation as:

$$S = I$$

This equation states that saving equals investment.

To understand the meaning of national saving, it is helpful to manipulate the definition a bit more. Let T denote the amount that the government collects from households in taxes (a withdrawal from the circular flow) minus the amount it pays back to households in the form of transfer payments (such as social security payments). We can then write national saving in either of two ways:

$$S = Y - C - G$$

or

$$S = (Y - T - C) + (T - G)$$

These equations are the same, because the two T s in the second equation cancel each other, but each reveals a different way of thinking about national saving. In particular, the second equation separates national saving into two pieces: private saving ($Y - T - C$) and public saving ($T - G$).

Consider each of these two pieces. **Private saving** is the amount of income that households have left after paying their taxes and paying for their consumption. In particular, because households receive income of Y , pay taxes of T and spend C on consumption, private saving is $Y - T - C$. **Public saving** is the amount of tax revenue that the government has left after paying for its spending. The government receives T in tax revenue and spends G on goods and services. If T exceeds G , the government runs a budget surplus because it receives more money than it spends. This surplus of $T - G$ represents public saving. If the government spends more than it receives in tax revenue, then G is larger than T . In this case, the government runs a budget deficit, and public saving $T - G$ is a negative number. In this case the government has to borrow money to fund spending by issuing sovereign debt in the form of bonds.

private saving the income that households have left after paying for taxes and consumption
public saving the tax revenue that the government has left after paying for its spending

Now consider how these accounting identities are related to financial markets. The equation $S = I$ reveals an important fact: for the economy as a whole, saving must be equal to investment. Yet this fact raises some important questions: what mechanisms lie behind this identity? What coordinates those people who are deciding how much to save and those people who are deciding how much to invest? The answer is the financial system. The bond market, the stock market, banks, investment funds, and other financial markets and intermediaries, stand between the two sides of the $S = I$ equation. They take in the nation's saving and direct it to the nation's investment.

The Meaning of Saving and Investment

The terms *saving* and *investment* can sometimes be confusing. Most people use these terms casually and sometimes interchangeably. In contrast, the macroeconomists who put together the national income accounts use these terms carefully and distinctly.

Consider an example. Suppose that Connah earns more than he spends and deposits his unspent income in a bank or uses it to buy a bond or some stock from a corporation. Because Connah's income exceeds his consumption, he adds to the nation's saving. Connah might think of himself as 'investing' his money, but a macroeconomist would call Connah's act saving rather than investment.

In the language of macroeconomics, investment refers to the purchase of new capital, such as equipment or buildings. When Gerry borrows from the bank to build himself a new house, he adds to the nation's investment. Similarly, when the O'Connell Corporation issues some new shares and uses the proceeds to build a new brass doorknocker factory, it also adds to the nation's investment.

Although the accounting identity $S = I$ shows that saving and investment are equal for the economy as a whole, this does not have to be true for every individual household or firm. Connah's saving can be greater than his investment, and he can deposit the excess in a bank. Gerry's saving can be less than his

investment, and he can borrow the shortfall from a bank. Banks and other financial institutions make these individual differences between saving and investment possible by allowing one person's saving to finance another person's investment.

SELF TEST Define private saving, public saving, national saving and investment. How are they related?

THE MARKET FOR LOANABLE FUNDS

Having discussed some of the important financial institutions in the economy and the macroeconomic role of these institutions, we are ready to build a model of financial markets. Our purpose in building this model is to explain how financial markets coordinate the economy's saving and investment. The model also gives us a tool with which we can analyse various government policies that influence saving and investment.

To keep things simple, we assume that the economy has only one financial market, called the **market for loanable funds**. All savers go to this market to deposit their saving, and all borrowers go to this market to get their loans. Thus, the term *loanable funds* refers to all income that people have chosen to save and lend out, rather than use for their own consumption. In the market for loanable funds, there is one interest rate, which is both the return to saving and the cost of borrowing.

market for loanable funds the market in which those who want to save supply funds and those who want to borrow to invest demand funds

The assumption of a single financial market, of course, is not literally true. As we have seen, the economy has many types of financial institutions. But, as we discussed in Chapter 2, the art in building an economic model is simplifying the world in order to explain it. For our purposes here, we can ignore the diversity of financial institutions and assume that the economy has a single financial market.

Supply and Demand for Loanable Funds

The economy's market for loanable funds, like other markets in the economy, is governed by supply and demand. To understand how the market for loanable funds operates, therefore, we first look at the sources of supply and demand in that market.

The supply of loanable funds comes from those people who have some extra income they want to save and lend out. This lending can occur directly, such as when a household buys a bond from a firm, or it can occur indirectly, such as when a household makes a deposit in a bank, which in turn uses the funds to make loans. In both cases, saving is the source of the supply of loanable funds.

The demand for loanable funds comes from households and firms who wish to borrow to make investments. This demand includes families taking out mortgages to buy homes. It also includes firms borrowing to buy new equipment or build factories. In both cases, investment is the source of the demand for loanable funds.

The interest rate is the price of a loan. It represents the amount that borrowers pay for loans and the amount that lenders receive on their saving. Because a high interest rate makes borrowing more expensive, the quantity of loanable funds demanded falls as the interest rate rises. Similarly, because a high interest rate makes saving more attractive, the quantity of loanable funds supplied rises as the interest rate rises. In other words, the demand curve for loanable funds slopes downwards, and the supply curve for loanable funds slopes upwards.

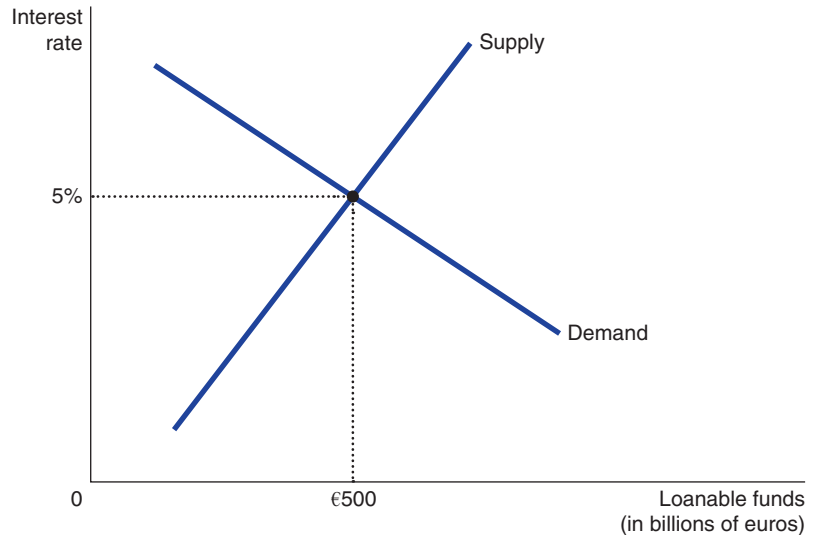
Figure 24.1 shows the interest rate that balances the supply and demand for loanable funds. In the equilibrium shown, the interest rate is 5 per cent, and the quantity of loanable funds demanded and the quantity of loanable funds supplied both equal €500 billion.

The adjustment of the interest rate to the equilibrium level occurs for the usual reasons. If the interest rate was lower than the equilibrium level, the quantity of loanable funds supplied would be less than the quantity of loanable funds demanded. The resulting shortage of loanable funds would encourage lenders

to raise the interest rate they charge. A higher interest rate would encourage saving (thereby increasing the quantity of loanable funds supplied) and discourage borrowing for investment (thereby decreasing the quantity of loanable funds demanded). Conversely, if the interest rate was higher than the equilibrium level, the quantity of loanable funds supplied would exceed the quantity of loanable funds demanded. As lenders competed for the scarce borrowers, interest rates would be driven down. In this way, the interest rate approaches the equilibrium level at which the supply and demand for loanable funds exactly balance.

FIGURE 24.1**The Market for Loanable Funds**

The interest rate in the economy adjusts to balance the supply and demand for loanable funds. The supply of loanable funds comes from national saving, including both private saving and public saving. The demand for loanable funds comes from firms and households that want to borrow for purposes of investment. Here the equilibrium interest rate is 5 per cent, and €500 billion of loanable funds are supplied and demanded.



Recall that economists distinguish between the real interest rate and the nominal interest rate. The nominal interest rate is the interest rate as usually reported – the monetary return to saving and cost of borrowing. The real interest rate is the nominal interest rate corrected for inflation; it equals the nominal interest rate minus the inflation rate. Because inflation erodes the value of money over time, the real interest rate more accurately reflects the real return to saving and cost of borrowing. Therefore, the supply and demand for loanable funds depends on the real (rather than nominal) interest rate, and the equilibrium in Figure 24.1 should be interpreted as determining the real interest rate in the economy. For the rest of this chapter, when you see the term *interest rate*, you should remember that we are talking about the real interest rate.

This model of the supply and demand for loanable funds shows that financial markets work much like other markets in the economy. Once we realize that saving represents the supply of loanable funds and investment represents the demand, we can see how the invisible hand coordinates saving and investment. When the interest rate adjusts to balance supply and demand in the market for loanable funds, it coordinates the behaviour of people who want to save (the suppliers of loanable funds) and the behaviour of people who want to invest (the demanders of loanable funds).

We can now use this analysis of the market for loanable funds to examine various government policies that affect the economy's saving and investment. Because this model is just supply and demand in a particular market, we analyse any policy using the three steps discussed in Chapter 3. First, we decide whether the policy shifts the supply curve or the demand curve. Second, we determine the direction of the shift. Third, we use the supply-and-demand diagram to see how the equilibrium changes.

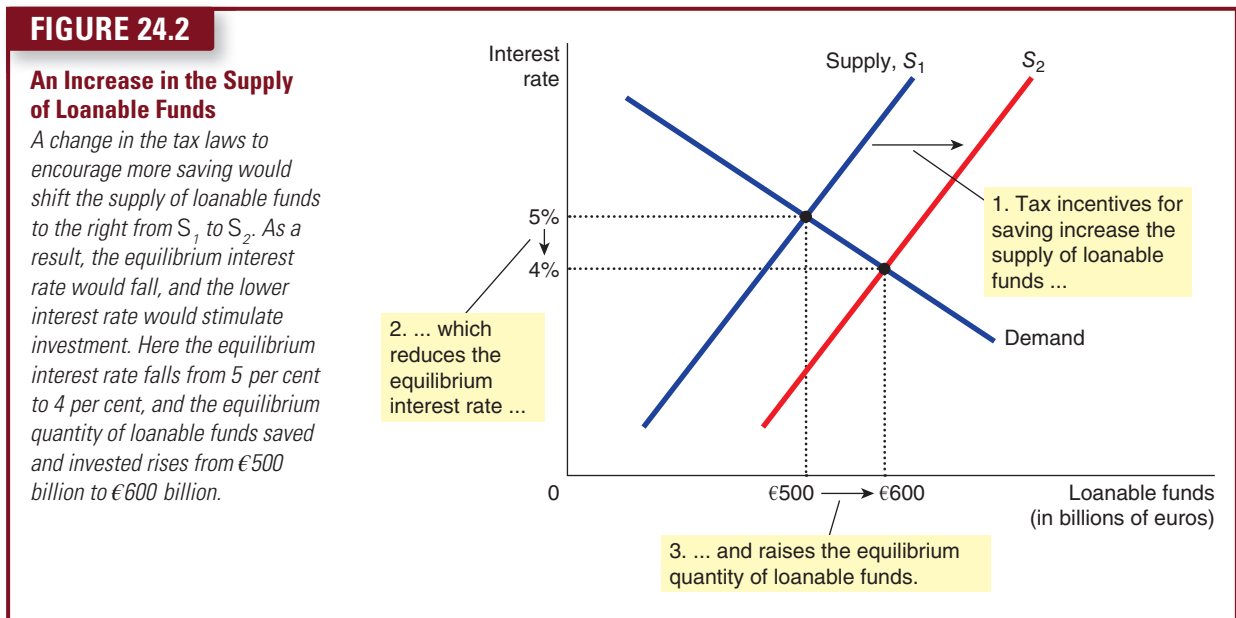
Policy 1: Saving Incentives

We know from Chapter 22 that saving is an important long-run determinant of a nation's productivity. Hence if a country can raise its saving rate, the growth rate of GDP should increase and, over time, the citizens of that country should enjoy a higher standard of living.

Many economists have used the principle that people respond to incentives to suggest that the savings rates in some countries are depressed because of tax laws that discourage saving. Governments collect revenue by taxing income, including interest and dividend income. To see the effects of this policy, consider a 25-year-old who saves €1,000 and buys a 30-year bond that pays an interest rate of 9 per cent. In the absence of taxes, the €1,000 grows to €13,268 when the individual reaches age 55. Yet if that interest is taxed at a rate of, say, 33 per cent, then the after-tax interest rate is only 6 per cent. In this case, the €1,000 grows to only €5,743 after 30 years. The tax on interest income substantially reduces the future pay-off from current saving and, as a result, reduces the incentive for people to save.

In response to this problem, many economists and some politicians have sometimes advocated replacing the current income tax with a consumption tax. Under a consumption tax, income that is saved would not be taxed until the saving is later spent; in essence, a consumption tax is like the value-added tax (VAT) that European countries impose on many goods and services. VAT is an indirect tax, however, levied on a good or service at the time it is purchased by a final consumer, whereas a consumption tax could also be a direct tax levied on an individual by calculating how much consumer expenditure they carried out over the year and taxing them on that, perhaps at higher and higher rates as the level of consumer expenditure rises.

A more modest proposal is to expand eligibility for special savings accounts that allow people to shelter some of their saving from taxation. Let's consider the effect of such a saving incentive on the market for loanable funds, as illustrated in Figure 24.2.



First, which curve would this policy affect? Because the tax change would alter the incentive for households to save *at any given interest rate*, it would affect the quantity of loanable funds supplied at each interest rate. Thus, the supply of loanable funds would shift. The demand for loanable funds would remain the same, because the tax change would not directly affect the amount that borrowers want to borrow at any given interest rate.

Second, which way would the supply curve shift? Because saving would be taxed less heavily, households would increase their saving by consuming a smaller fraction of their income. Households would use this additional saving to increase their deposits in banks or to buy more bonds. The supply of loanable funds would increase, and the supply curve would shift to the right from S_1 to S_2 , as shown in Figure 24.2.

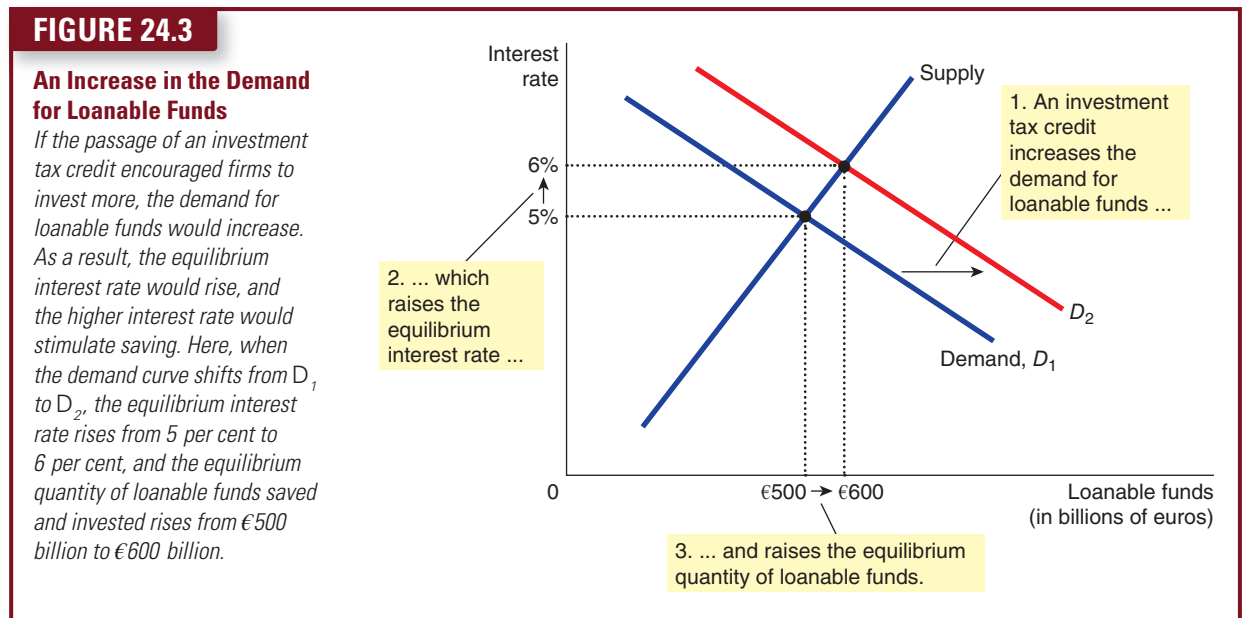
Finally, we can compare the old and new equilibria. In the figure, the increased supply of loanable funds reduces the interest rate from 5 per cent to 4 per cent. The lower interest rate raises the quantity of loanable funds demanded from €500 billion to €600 billion. That is, the shift in the supply curve moves

the market equilibrium along the demand curve. With a lower cost of borrowing, households and firms are motivated to borrow more to finance greater investment. Thus, if a reform of the tax laws encouraged greater saving, the result would be lower interest rates and greater investment.

Although this analysis of the effects of increased saving is widely accepted among economists, there is less consensus about what kinds of tax changes should be enacted. Many economists endorse tax reform aimed at increasing saving in order to stimulate investment and growth. Yet others are sceptical that these tax changes would have much effect on national saving. These sceptics also doubt the equity of the proposed reforms. They argue that, in many cases, the benefits of the tax changes would accrue primarily to the wealthy, who are least in need of tax relief.

Policy 2: Investment Incentives

Suppose that the government passed a tax reform aimed at making investment more attractive. In essence, this is what the government does when it institutes an *investment tax credit*, which it does from time to time. An investment tax credit gives a tax advantage to any firm building a new factory or buying a new piece of equipment. Let's consider the effect of such a tax reform on the market for loanable funds, as illustrated in Figure 24.3.



First, would the reform affect supply or demand? Because the tax credit would reward firms that borrow and invest in new capital, it would alter investment at any given interest rate and, thereby, change the demand for loanable funds. By contrast, because the tax credit would not affect the amount that households save at any given interest rate, it would not affect the supply of loanable funds.

Second, which way would the demand curve shift? Because firms would have an incentive to increase investment at any interest rate, the quantity of loanable funds demanded would be higher at any given interest rate. Thus, the demand curve for loanable funds would move to the right, as shown by the shift from D_1 to D_2 in the figure.

Third, consider how the equilibrium would change. In Figure 24.3, the increased demand for loanable funds raises the interest rate from 5 per cent to 6 per cent, and the higher interest rate in turn increases the quantity of loanable funds supplied from €500 billion to €600 billion, as households respond by increasing the amount they save. This change in household behaviour is represented here as a movement along the supply curve. Thus, if a reform of the tax system encouraged greater investment, the result would be higher interest rates and greater saving.

Policy 3: Government Budget Deficits and Surpluses

A perpetual topic of political debate is the status of the government budget. Recall that a *budget deficit* is an excess of government spending over tax revenue. Governments finance budget deficits by borrowing in the bond market, and the accumulation of past government borrowing is called the *government debt*. A *budget surplus*, an excess of tax revenue over government spending, can be used to repay some of the government debt. If government spending exactly equals tax revenue, the government is said to have a *balanced budget*.

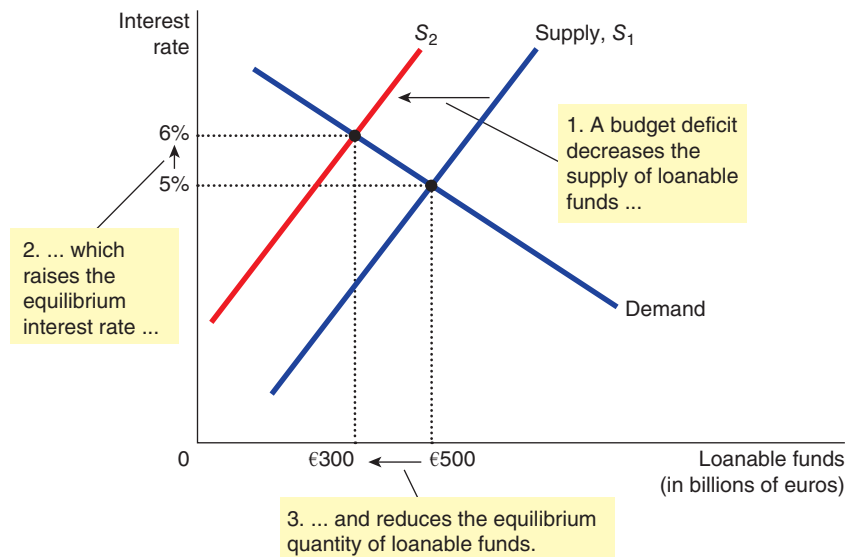
Imagine that the government starts with a balanced budget and then, because of a tax cut or a spending increase, starts running a budget deficit. We can analyse the effects of the budget deficit by following our three steps in the market for loanable funds, as illustrated in Figure 24.4.

First, which curve shifts when the government starts running a budget deficit? Recall that national saving – the source of the supply of loanable funds – is composed of private saving and public saving. A change in the government budget balance represents a change in public saving and, thereby, in the supply of loanable funds. Because the budget deficit does not influence the amount that households and firms want to borrow to finance investment at any given interest rate, it does not alter the demand for loanable funds.

FIGURE 24.4

The Effect of a Government Budget Deficit

When the government spends more than it receives in tax revenue, the resulting budget deficit lowers national saving. The supply of loanable funds decreases, and the equilibrium interest rate rises. Thus, when the government borrows to finance its budget deficit, it crowds out households and firms who otherwise would borrow to finance investment. Here, when the supply shifts from S_1 to S_2 , the equilibrium interest rate rises from 5 per cent to 6 per cent, and the equilibrium quantity of loanable funds saved and invested falls from €500 billion to €300 billion.



Second, which way does the supply curve shift? When the government runs a budget deficit, public saving is negative, and this reduces national saving. In other words, when the government borrows to finance its budget deficit, it reduces the supply of loanable funds available to finance investment by households and firms. Thus, a budget deficit shifts the supply curve for loanable funds to the left from S_1 to S_2 , as shown in Figure 24.4.

Third, we can compare the old and new equilibria. In the figure, when the budget deficit reduces the supply of loanable funds, the interest rate rises from 5 per cent to 6 per cent. This higher interest rate then alters the behaviour of the households and firms that participate in the loan market. In particular, many demanders of loanable funds are discouraged by the higher interest rate. Fewer families buy

new homes, and fewer firms choose to build new factories. The fall in investment because of government borrowing is called **crowding out** and is represented in the figure by the movement along the demand curve from a quantity of €500 billion in loanable funds to a quantity of €300 billion. That is, when the government borrows to finance its budget deficit, it crowds out private borrowers who are trying to finance investment.

crowding out a decrease in investment that results from government borrowing

Thus, the most basic lesson about budget deficits follows directly from their effects on the supply and demand for loanable funds: when the government reduces national saving by running a budget deficit, the interest rate rises and investment falls. Because investment is important for long-run economic growth, government budget deficits reduce the economy's growth rate.

Government budget surpluses work just the opposite way to budget deficits. When government collects more in tax revenue than it spends, it saves the difference by retiring some of the outstanding government debt. This budget surplus, or public saving, contributes to national saving. Thus, a budget surplus increases the supply of loanable funds, reduces the interest rate and stimulates investment. Higher investment, in turn, means greater capital accumulation and more rapid economic growth.

CONCLUSION

'Neither a borrower nor a lender be', Polonius advises his son in Shakespeare's *Hamlet*. If everyone followed this advice, this chapter would have been unnecessary.

Few economists would agree with Polonius. In our economy, people borrow and lend often, and usually for good reason. You may borrow one day to start your own business or to buy a home. And people may lend to you in the hope that the interest you pay will allow them to enjoy a more prosperous retirement. The financial system has the job of coordinating all this borrowing and lending activity.

In many ways, financial markets are like other markets in the economy. The price of loanable funds – the interest rate – is governed by the forces of supply and demand, just as other prices in the economy are. And we can analyse shifts in supply or demand in financial markets as we do in other markets. One of the *Ten Principles of Economics* is that markets are usually a good way to organize economic activity. This principle applies to financial markets as well. When financial markets bring the supply and demand for loanable funds into balance, they help allocate the economy's scarce resources to their most efficient use.

In one way, however, financial markets are special. Financial markets, unlike most other markets, serve the important role of linking the present and the future. Those who supply loanable funds – savers – do so because they want to convert some of their current income into future purchasing power. Those who demand loanable funds – borrowers – do so because they want to invest today in order to have additional capital in the future to produce goods and services. Thus, well-functioning financial markets are important not only for current generations but also for future generations who will inherit many of the resulting benefits.

IN THE NEWS



Borrowing in Developing Countries

The growth of developing countries has been the subject of a large amount of interest amongst economists and politicians. These countries may provide the markets for the developed countries in the future but the speed at which they develop may be influenced by a number of factors as outlined in this article.

The BNDES and Crowding Out

Brazil is one of the so-called BRIC countries – countries which are developing strongly and include Russia, China and India in addition to Brazil. In 1952, the Brazilian government set up a development bank called Banco Nacional de Desenvolvimento Econômico e Social (BNDES) as a vehicle to help with the development of the country. As Brazil has grown, the BNDES has become more powerful and a bigger player in the capital markets, so much so that some critics are now claiming that it is taking too big a role in borrowing for investment and is crowding out long-term lending of the sort required for key infrastructure projects by private-sector banks in Brazil.

To put the potential problem into context, the BNDES lent almost \$47 billion (€35 billion) more than the World Bank in 2009–2010 to Brazilian corporations. It lent \$76.3 billion (€57 billion) in 2012, 12 per cent up on

the previous year and was the major investor in a number of share offerings, one of the most notable being a \$70.1 billion (€52.5 billion) share offering by Brazil's oil company, Petrobras.

The sheer size of the involvement of the BNDES in investment in Brazil has raised concerns that private investment is being crowded out. When corporations borrow from the BNDES, they have to pay the Taxa de Juros de Longo Prazo (TJLP), a long-term interest rate set by the Brazilian central bank's monetary policy committee. This rate is lower than the rate at which corporations can borrow from the commercial banking sector by, in some cases, 2.5 per cent. In other cases, firms can borrow for capital investment at even lower rates which, when inflation is taken into account (which in early 2013 was running at over 6 per cent), means that corporations can borrow at negative interest rates. Corporations, therefore, are able to access loans

for investment at subsidized rates. With that sort of incentive, the opportunity for commercial lenders to be able to compete is limited, given that the BNDES accounts for over 75 per cent of long-term loans (defined as in excess of three years). When the BNDES is so dominant in the market for funds, it is likely to be a barrier to Brazil's future ongoing development if a thriving private sector capital market such as those which operate in developed countries is not allowed to grow. For those corporations which are not able to access BNDES funding, the additional cost of securing private lending is much higher and as a result they are less competitive.

Questions

- 1 **Why is the development of infrastructure projects so vital for a developing country like Brazil?**
- 2 **In what sense is the involvement of the BNDES crowding out investment from private lenders?**
- 3 **What do you think might be the advantages and disadvantages of the subsidized interest rates offered by the BNDES to Brazilian corporations and the economy as a whole?**
- 4 **In the context of corporations borrowing from the BNDES and from other private sector sources what is the relevance of the real interest rate?**
- 5 **Why might the continued involvement of the BNDES in lending in Brazil be a barrier to the country's growth prospects in the future?**

An offshore oil rig in the Campos Basin in Rio de Janeiro, state owned by Petrobras, one of the companies raising capital for expansion from BNDES.



SUMMARY

- The financial system of an advanced economy is made up of many types of financial institutions, such as the bond market, the stock market, banks and investment funds. All of these institutions act to direct the resources of households, who want to save some of their income, into the hands of households and firms who want to borrow.
- National income accounting identities reveal some important relationships among macroeconomic variables. In particular, for a closed economy, national saving must equal investment. Financial institutions are the mechanism through which the economy matches one person's saving with another person's investment.
- The interest rate is determined by the supply and demand for loanable funds. The supply of loanable funds comes from households who want to save some of their income and lend it out. The demand for loanable funds comes from households and firms who want to borrow for investment. To analyse how any policy or event affects the interest rate, one must consider how it affects the supply and demand for loanable funds.
- National saving equals private saving plus public saving. A government budget deficit represents negative public saving and, therefore, reduces national saving and the supply of loanable funds available to finance investment. When a government budget deficit crowds out investment, it reduces the growth of productivity and GDP.

QUESTIONS FOR REVIEW

- 1 What is the role of the financial system?
- 2 Name and describe two markets that are part of the financial system in the economy of the country in which you are studying.
- 3 What is a financial intermediary? Name and describe two financial intermediaries.
- 4 Why is it important for people who own stocks and bonds to diversify their holdings? What type of financial institution makes diversification easier?
- 5 What is national saving? What is private saving? What is public saving? How are these three variables related?
- 6 How do credit default swaps help reduce the risk involved in buying bonds?
- 7 What is a collateralized debt obligation?
- 8 What is investment? How is it related to national saving?
- 9 Describe a change in the tax system that might increase private saving. If this policy were implemented, how would it affect the market for loanable funds?
- 10 What is a government budget deficit? How does it affect interest rates, investment and economic growth?

PROBLEMS AND APPLICATIONS

- 1 For each of the following pairs, which bond would you expect to pay a higher interest rate? Explain.
 - a. A bond of the UK government or a bond of an east European government.
 - b. A bond that repays the principal in year 2020 or a bond that repays the principal in year 2035.
 - c. A bond from BP or a bond from a software company operating from a business park on the outskirts of a large city.
 - d. A bond issued by the national government or a bond issued by a local authority.
- 2 What is the difference between gambling at cards or in lotteries or on the race track and gambling on the stock market? What social purpose do you think is served by the existence of the stock market?
- 3 Declines in share prices are sometimes viewed as harbingers of future declines in real GDP. Why do you suppose that might be true?
- 4 When the Greek government asked for support from the European Union to repay bonds that were coming up for maturity between 2008 and 2010, interest rates rose on bonds issued by a number of other EU countries like Portugal, Spain and Ireland also rose. Why do you suppose this happened? What can you predict happened to the price of bonds from these countries during this period?
- 5 Many workers hold large amounts of stock issued by the firms at which they work. Why do you suppose companies encourage this behaviour? Why might a person not want to hold stock in the company where he works?

- 6** Explain the difference between saving and investment as defined by a macroeconomist. Which of the following situations represent investment? Saving? Explain.
- Your family takes out a mortgage and buys a new house.
 - You use your €500 wage payment to buy stock in BP.
 - Your flatmate earns €200 and deposits it in her account at a bank.
 - You borrow €5,000 from a bank to buy a car to use in your pizza delivery business.
- 7** Suppose GDP is €5 trillion, taxes are €1.5 trillion, private saving is €0.5 trillion and public saving is €0.2 trillion. Assuming this economy is closed, calculate consumption, government purchases, national saving and investment.
- 8** Suppose that BP is considering exploring a new oil field.
- Assuming that BP needs to borrow money in the bond market to finance the purchase of new oil rigs and drilling machinery, why would an increase in interest rates affect BP's decision about whether to carry out the exploration?
 - If BP has enough of its own funds to finance the development of the new oil field without borrowing, would an increase in interest rates still affect BP's decision about whether to undertake the new project? Explain.
- 9** Suppose the government borrows €5 billion more next year than this year.
- Use a supply-and-demand diagram to analyse this policy. Does the interest rate rise or fall?
 - What happens to investment? To private saving? To public saving? To national saving? Compare the size of the changes to the €5 billion of extra government borrowing.
 - How does the elasticity of supply of loanable funds affect the size of these changes?
 - How does the elasticity of demand for loanable funds affect the size of these changes?
 - Suppose households believe that greater government borrowing today implies higher taxes to pay off the government debt in the future. What does this belief do to private saving and the supply of loanable funds today? Does it increase or decrease the effects you discussed in parts (a) and (b)?
- 10** This chapter explains that investment can be increased both by reducing taxes on private saving and by reducing the government budget deficit.
- Why is it difficult to implement both of these policies at the same time?
 - What would you need to know about private saving in order to judge which of these two policies would be a more effective way to raise investment?

25

THE BASIC TOOLS OF FINANCE

You will almost certainly have had to deal with the economy's financial system already in your life and this will only increase as you leave education and move into work. You will deposit your savings in a bank account, or you might take out a loan to buy a car or a mortgage to buy a house. Once you take a job, you will have to decide whether to invest your retirement account in company shares, bonds or other financial instruments (or you may buy a personal pension and let the pension company do this for you). You may try to put together your own share portfolio, and then you will have to decide between betting on established companies or newer ones. You will have to take out assurance and insurance policies to cover various risks such as death, fire and theft to property, vehicle insurance and so on. And whenever you watch the evening news, you will hear reports about whether the stock market is up or down, together with the attempts to explain why the market behaves as it does.

If you reflect for a moment on the many financial decisions you will make through your life, you will see two related elements in almost all of them – time and risk. As we saw in the preceding chapter, the financial system coordinates the economy's saving and investment. Thus, it concerns decisions we make today that will affect our lives in the future. But the future is unknown. When a person decides to allocate some saving, or a firm decides to undertake an investment, the decision is based on a guess (albeit an often informed guess) about the likely future result – but the actual result could end up being very different.

This chapter introduces some tools that help us understand the decisions that people make as they participate in financial markets. The field of **finance** – a sub-discipline of economics – develops these tools in great detail, and you may choose to take courses that focus on this topic. But because the financial system is so important to the functioning of the economy, many of the basic insights of finance are central to understanding how the economy works. The tools of finance may also help you think through some of the decisions that you will make in your own life.

finance the field of economics that studies how people make decisions regarding the allocation of resources over time and the handling of risk

This chapter takes up three topics. First, we discuss how to compare sums of money at different points in time. Second, we discuss how to manage risk. Third, we build on our analysis of time and risk to examine what determines the value of an asset, such as a share of stock.

PRESENT VALUE: MEASURING THE TIME VALUE OF MONEY

Imagine that someone offered to give you €100 today or €100 in ten years. Which would you choose? The answer to this question should be easy. Getting €100 today is better, because you can always deposit the money in a bank, still have it in ten years and earn interest on the €100 along the way. The lesson: money today is more valuable than the same amount of money in the future.

Now consider a harder question: imagine that someone offered you €100 today or €200 in ten years. Which would you choose? To answer this question, you need some way to compare sums of money from

different points in time. Economists do this with a concept called *present value*. The **present value** of any future sum of money is the amount today that would be needed, at current interest rates, to produce that future sum.

present value the amount of money today that would be needed to produce, using prevailing interest rates, a given future amount of money

To learn how to use the concept of present value, let's work through a couple of simple examples.

Question: If you put €100 in a bank account today, how much will it be worth in N years? That is, what will be the **future value** of this €100?

future value the amount of money in the future that an amount of money today will yield, given prevailing interest rates

Answer: Let's use r to denote the interest rate expressed in decimal form (so an interest rate of 5 per cent means $r = 0.05$). Suppose that interest is paid annually and that the interest paid remains in the bank account to earn more interest – a process called **compounding**. Then the €100 will become:

$$\begin{array}{ll} (1+r) \times \text{€}100 & \text{after one year} \\ (1+r) \times (1+r)\text{€}100 & \text{after two years} \\ (1+r) \times (1+r)(1+r)\text{€}100 & \text{after three years} \\ (1+r)^N \times \text{€}100 & \text{after } N \text{ years} \end{array}$$

For example, if we are investing at an interest rate of 5 per cent for ten years, then the future value of the €100 will be $(1.05)^{10} \times \text{€}100$, which is €163 (rounded up to the nearest whole euro. Note: we will follow this rounding up for the remainder of the chapter).

compounding the accumulation of a sum of money in, say, a bank account, where the interest earned remains in the account to earn additional interest in the future

Question: Now suppose you are going to be paid €200 in N years. What is the *present value* of this future payment? That is, how much would you have to deposit in a bank right now to yield €200 in N years?

Answer: To answer this question, just turn the previous answer on its head. In the first question, we computed a future value from a present value by *multiplying* by the factor $(1+r)^N$. To compute a present value from a future value, we *divide* by the factor $(1+r)^N$. Thus, the present value of €200 in N years is $\text{€}200/(1+r)^N$. If that amount is deposited in a bank today, after N years it would become $(1+r)^N \times [\text{€}200/(1+r)^N]$, which is €200. For instance, if the interest rate is 5 per cent, the present value of €200 in ten years is $\text{€}200/(1.05)^{10}$, which is €123.

This illustrates the general formula: if r is the interest rate, then an amount X to be received in N years has present value of $X/(1+r)^N$.

Let's now return to our earlier question: should you choose €100 today or €200 in ten years? We can infer from our calculation of present value that if the interest rate is 5 per cent, you should prefer the €200 in ten years. The future €200 has a present value of €123, which is greater than €100. You are better off waiting for the future sum.

Notice that the answer to our question depends on the interest rate. If the interest rate were 8 per cent, then the €200 in ten years would have a present value of $\text{€}200/(1.08)^{10}$, which is only €93. In this case, you should take the €100 today. Why should the interest rate matter for your choice? The answer is that the higher the interest rate, the more you can earn by depositing your money at the bank, so the more attractive getting €100 today becomes.

Applying the Concept of Present Value

The concept of present value is useful in many applications, including the decisions that companies face when evaluating investment projects. For instance, imagine that Citroën is thinking about building a new car factory. Suppose that the factory will cost €100 million today and will yield the company €200 million in ten years. Should Citroën undertake the project? You can see that this decision is exactly like the one we have been studying. To make its decision, the company will compare the present value of the €200 million return to the €100 million cost.

The company's decision, therefore, will depend on the interest rate. If the interest rate is 5 per cent, then the present value of the €200 million return from the factory is €123 million, and the company will choose to pay the €100 million cost. By contrast, if the interest rate is 8 per cent, then the present value of the return is only €93 million, and the company might decide to forgo the project. Thus, the concept of present value helps explain why investment – and thus the quantity of loanable funds demanded – declines when the interest rate rises.

Here is another application of present value: suppose you win the Euro Millions Lottery and you are given a choice between €20,000 a year for 50 years (totalling €1,000,000) or an immediate payment of €400,000. Which would you choose? To make the right choice, you need to calculate the present value of the stream of payments. After performing 50 calculations similar to those above (one calculation for each payment) and adding up the results, you would learn that the present value of this million-euro prize at a 7 per cent interest rate is only €256,000. You are better off picking the immediate payment of €400,000. The million euros may seem like more money, but the future cash flows, once discounted to the present, are worth far less.

SELF TEST The interest rate is 7 per cent. What is the present value of €150 to be received in ten years?

MANAGING RISK

Life is full of gambles. When you go skiing, you risk breaking your leg in a fall. When you cycle to work or university, you risk being knocked off your bike by a car. When you put some of your savings in the stock market, you risk a fall in prices. **Risk** is the probability of something happening which results in a loss or some degree of hazard or damage. The rational response to risk is not necessarily to avoid it at any cost, but to take it into account in your decision-making. Let's consider how a person might do that.

risk the probability of something happening which results in a loss or some degree of hazard or damage

Risk Aversion

Most people are **risk averse**. This means more than simply people dislike bad things happening to them. It means that they dislike bad things more than they like comparable good things. (This is also reflected in *loss aversion* – research suggests that losing something makes people twice as miserable as gaining something makes them happy!)

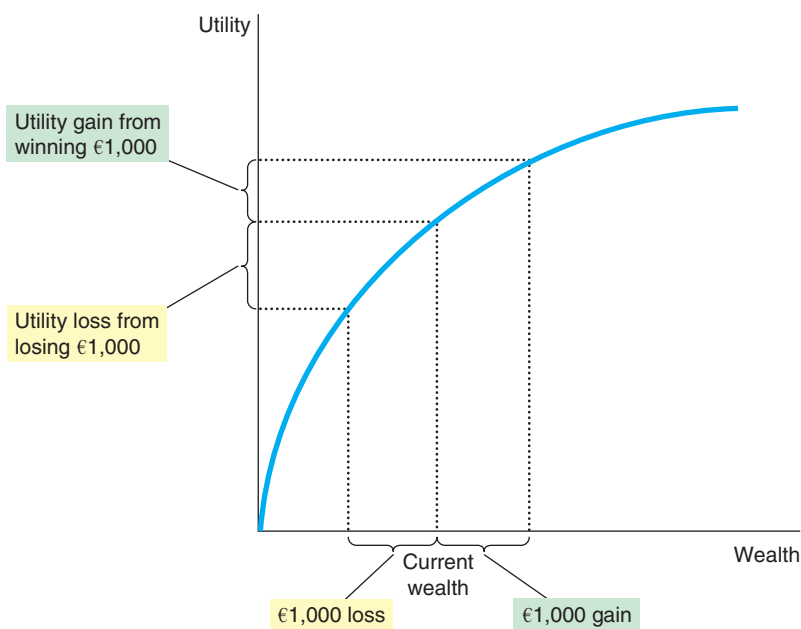
risk averse exhibiting a dislike of uncertainty

For example, suppose a friend offers you the following opportunity. He will flip a coin. If it comes up heads, he will pay you €1,000. But if it comes up tails, you will have to pay him €1,000. Would you accept the bargain? You wouldn't if you were risk averse, even though the probability of winning is the same as the probability of losing (50 per cent). For a risk-averse person, the pain from losing the €1,000 would exceed the gain from winning €1,000.

Economists have developed models of risk aversion using the concept of *utility*, which is a person's subjective measure of well-being or satisfaction. Every level of wealth provides a certain amount of utility, as shown by the utility function in Figure 25.1. But the function exhibits the property of diminishing marginal utility: the more wealth a person has, the less utility he gets from an additional euro. Thus, in the figure, the utility function gets flatter as wealth increases. Because of diminishing marginal utility, the utility lost from losing the €1,000 bet is more than the utility gained from winning it. As a result, people are risk averse.

FIGURE 25.1**The Utility Function**

This utility function shows how utility, a subjective measure of satisfaction, depends on wealth. As wealth rises, the utility function becomes flatter, reflecting the property of diminishing marginal utility. Because of diminishing marginal utility, a €1,000 loss decreases utility by more than a €1,000 gain increases it.



Risk aversion provides the starting point for explaining various things we observe in the economy. Let's consider three of them: insurance, diversification and the risk–return trade-off.

The Markets for Insurance

One way to deal with risk is to buy insurance. The general feature of insurance contracts is that a person facing a risk pays a fee to an insurance company, which in return agrees to accept all or part of the risk. There are many types of insurance. Car insurance covers the risk of your being in a car accident, fire insurance covers the risk that your house will burn down, and life assurance covers the risk that you will die and leave your family without your income. There is also insurance against the risk of living too long and running out of money: for a fee paid today, an insurance company will pay you an *annuity* – a regular income every year until you die (otherwise known as a pension).

In a sense, every insurance contract is a gamble. It is possible that you will not be in a car accident or that your house will not burn down. In most years, you will pay the insurance company the premium and get nothing in return except peace of mind. Indeed, the insurance company is counting on the fact that most people will not make claims on their policies; otherwise, it couldn't pay out the large claims to those few who are unlucky and still stay in business.

From the standpoint of the economy as a whole, the role of insurance is not to eliminate the risks inherent in life but to spread them around more efficiently. Consider fire insurance, for instance. Owning fire insurance does not reduce the risk of losing your home in a fire. But, if that unlucky event occurs, the insurance company compensates you. The risk, rather than being borne by you alone, is shared among the thousands of insurance company shareholders. Because people are risk averse, it is easier for 10,000 people to bear 1/10,000 of the risk than for one person to bear the entire risk themselves.

The markets for insurance suffer from two types of problems that impede their ability to spread risk. One problem is *adverse selection*: a high-risk person is more likely to apply for insurance than a low-risk person. A second problem is *moral hazard*: after people buy insurance, they have less incentive to be careful about their risky behaviour. Insurance companies are aware of these problems, and the price of insurance reflects the actual risks that the insurance company will face after the insurance is bought. The high price of insurance is why some people, especially those who know themselves to be low risk, decide against buying insurance and, instead, endure some of life's uncertainty on their own.

FYI



Pricing Risk

We have seen how bond issues are a means by which firms can borrow money. The buyer has to have confidence that she will get her money back and also receive an appropriate reward for lending the money in the first place. There is a risk involved that the issuer will not be able to pay back the money and that risk is associated with a probability. If the issuer is very sound then the probability of default may be close to zero, but if the issuer is extremely weak then the probability is closer to 1. We saw in Chapter 24 how financial markets now deal in pools of debt (collections of different types of loans sold to an investor). As debt is pooled the outcomes become more varied. In any given pool of mortgage debt, for example, there will be some borrowers who will default and not be able to pay off their mortgages – possibly because of family bereavement or loss of their job. Other mortgage holders may look to pay off their mortgages early, some may move house and thus settle their mortgage, some will increase monthly payments or pay lump sums to help reduce the repayment period of their mortgage and so on. Assessing probabilities

with such a wide range of outcomes becomes difficult.

The risk involved with such debt is therefore difficult to assess with any certainty. However, investors want to price risk as part of their decision-making so that they can judge the value of an asset. If an asset is very risky then the returns expected will be higher and vice versa. In order to have an efficient market, that risk has to be priced and the information on which the risk is based has to be reliable, accurate – and understood.

Let's consider an example. In your class there may be a number of students that you associate with every day. Take any one individual and we can identify a number of 'risks' for that person. For example, there is a risk that the individual:

- fails his exams and has to leave the course
- will be involved in a car crash
- will travel on an aircraft more than five times a year
- may be mugged
- may get flu.

What are the chances of these events happening? The analysis of such outcomes is what actuaries in

the insurance industry have to do. An estimate of the probability of such events happening can be derived from analysis of data, specifically historical data. It is possible, therefore, to gather data on the average 19-year-old student coming from a particular area and with a particular background, and use this data to arrive at the probability of the event occurring. Historical data tell us, for example that young people aged between 18 and 24 are more likely to be mugged than the elderly, despite popular perception. If we are able to identify these probabilities then they can be priced. Securities can be issued based on the chances of these things happening – the more likely the event to occur, the higher the price and vice versa.

Whilst it may be possible to identify probabilities for an individual it may be more problematic when looking at relationships between individuals. For example, if person X fails her exams what is the probability that you will also fail your exams? If that individual gets flu what are the chances that you also get flu? In both cases the probability might depend on your

relationship with that person. If you spend a lot of time with that person then it may be that you share similar distractions – going out every night instead of studying, skipping lectures to play pool and so on. If this were the case then the probability of you also failing your exams and getting flu might be high, but if you have no relationship at all then the chance of you sharing bad habits which lead you to also failing your exams are lower. However, given that you share some time with that person in a lecture hall or seminar room, for example, might mean that the probability of also getting swine flu is relatively high.

Looking at such relationships involves the concept of correlation – the extent to which there is any statistical relationship between two variables. If person X is involved in a car crash (and you were not in the car with him) what is the chance of you also being involved in a car crash? The chances are the correlation is very low; the probability of you both getting mugged is

higher regardless of the relationship between you and so there will be a stronger correlation in this instance. The correlation is likely to become more and more unstable the more variables are introduced (number of students in this example). In the case of pools of debt, the same problems arise and the efficiency of the information on which investors are basing their decision becomes ever more complex; probabilities become very difficult to assess and therefore to price.

Actuaries have been studying these types of correlations for some years. Issuing life assurance involves a risk. Life assurance means that the event covered – the death of an individual – will occur at some point in the future (unlike insurance where the event might never happen). The job of the actuary is to provide information to the insurer on the chances of death occurring under different situations. Where information becomes available which indicates risk factors change, actuaries

have to incorporate this into models to help insurers price the risk adequately (i.e. set the premiums for the policy). In the 1980s much work was done on studying a phenomenon known as stress cardiomyopathy, a condition for which, in the wake of some exceptional emotional trauma, the human brain releases chemicals into the bloodstream that leads to a weakening of the heart and an increased chance of death. The condition has been referred to as a 'broken heart' because it seemed to manifest itself in particular where one partner died soon after the death of the other. A study by Spreeuw, J. and Wang, X. (2008), '*Modelling the short-term dependence between two remaining lifetimes*', showed that following the death of a female partner, a male was over six times more likely to die than normal and women more than twice as likely. The conclusion for the insurance industry is to take such information and build it into the pricing of offering joint-life policies.

Diversification of Idiosyncratic Risk

In 2009 thousands of investors lost their savings as a result of fraud committed by former stock broker Bernard Madoff in the United States. The fraud, known as a Ponzi scheme, involved Madoff taking money from investors on the promise of high returns through a supposedly sophisticated trading and hedging strategy with little apparent risk to the investor. Madoff did not invest the money he received but instead relied on attracting new investors, whose money was used to pay existing investors and hence maintain the fraud. Madoff eventually received a prison sentence of 150 years for crimes such as serious fraud, lying to federal securities regulators and money laundering. The saddest part of the story, however, involved the thousands of investors. The estimated value of the fraud was some \$65 billion. Some of the victims were wealthy film stars but there were also many who had saved hard for their future and charities, all of whom lost their life savings. At least two suicides have been linked with the fraud.

If there is one piece of practical advice that finance offers to risk-averse people, it is this: 'Don't put all your eggs in one basket,' You may have heard this before, but finance has turned this traditional wisdom into a science. It goes by the name **diversification**.

diversification the reduction of risk achieved by replacing a single risk with a large number of smaller unrelated risks

The market for insurance is one example of diversification. Imagine a town with 10,000 homeowners, each facing the risk of a house fire. If someone starts an insurance company and each person in town becomes both a shareholder and a policyholder of the company, they all reduce their risk through diversification. Each person now faces 1/10,000 of the risk of 10,000 possible fires, rather than the entire risk of a single fire in his own home. Unless the entire town catches fire at the same time, the downside that each person faces is much smaller.

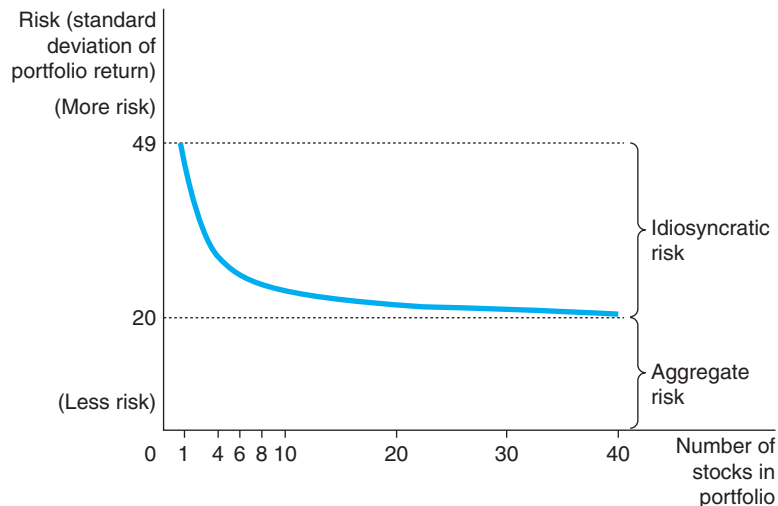
When people use their savings to buy financial assets, they can also reduce risk through diversification. A person who buys stock in a company is placing a bet on the future profitability of that company. That bet is often quite risky because companies' fortunes are hard to predict. Microsoft evolved from a start-up by some geeky teenagers into one of the world's most valuable companies in only a few years; Enron, a former US energy company which became insolvent with the loss of millions to investors and the jobs of workers because of fraud, went from one of the world's most respected companies to an almost worthless one in only a few months. Fortunately, a shareholder need not tie his own fortune to that of any single company. Risk can be reduced by placing a large number of small bets, rather than a small number of large ones.

Figure 25.2 shows how the risk of a portfolio of shares depends on the number of shares in the portfolio. Risk is measured here with a statistic called *standard deviation*, which you may have learned about in a mathematics or statistics course. Standard deviation measures the volatility of a variable – that is, how much the variable is likely to fluctuate. The higher the standard deviation of a portfolio's return, the more volatile and riskier it is.

FIGURE 25.2

Diversification Reduces Risk

This figure shows how the risk of a portfolio, measured here with a statistic called standard deviation (a measure of the variability about the mean), depends on the number of shares in the portfolio. The investor is assumed to put an equal percentage of his portfolio in each of the shares. Increasing the number of shares reduces the amount of risk in a stock portfolio, but it does not eliminate it.



The figure shows that the risk of a stock portfolio reduces substantially as the number of shares increase. For a portfolio with a single share, the standard deviation is 49 per cent. Going from 1 share to 10 shares eliminates about half the risk. Going from 10 to 20 shares reduces the risk by another 13 per cent. As the number of shares continues to increase, risk continues to fall, although the reductions in risk after 20 or 30 shares are small.

Notice that it is impossible to eliminate all risk by increasing the number of stocks or shares in the portfolio. Diversification can eliminate **idiosyncratic risk** – the uncertainty associated with the specific

companies. But diversification cannot eliminate **aggregate risk** – the uncertainty associated with the entire economy, which affects all companies. For example, when the economy goes into a recession, most companies experience falling sales, reduced profit and low stock returns. Diversification reduces the risk of holding stocks, but it does not eliminate it.

idiosyncratic risk risk that affects only a single economic actor
aggregate risk risk that affects all economic actors at once

The Trade-Off Between Risk and Return

One of the *Ten Principles of Economics* is that people face trade-offs. The trade-off that is most relevant for understanding financial decisions is the trade-off between risk and return.

As we have seen, there are risks inherent in holding shares, even in a diversified portfolio. But risk-averse people are willing to accept this uncertainty because they are compensated for doing so. Historically, shares have offered much higher rates of return than alternative financial assets, such as bonds and bank savings accounts. Over the past two centuries, stocks have offered an average real return of about 8 per cent per year, while short-term government bonds paid a real return of only 3 per cent per year.

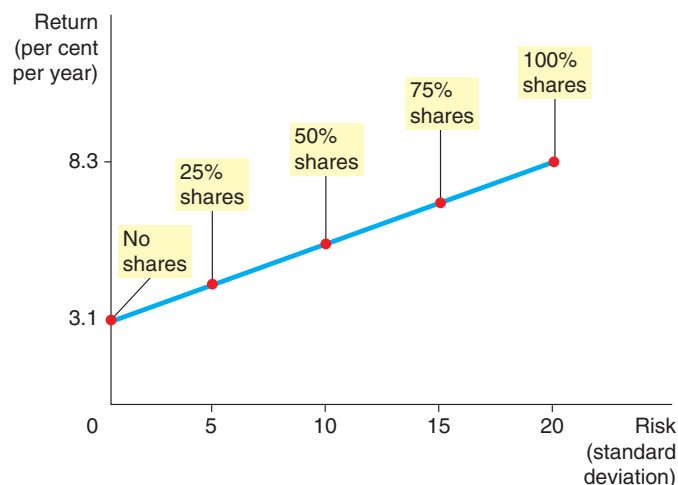
When deciding how to allocate their savings, people have to decide how much risk they are willing to undertake to earn the higher return. Figure 25.3 illustrates the risk-return trade-off for a person choosing to allocate his portfolio between two asset classes:

- The first asset class is a diversified group of risky stocks, with an average return of 8 per cent and a standard deviation of 20 per cent. (You may recall from mathematics or statistics courses that a normal random variable stays within two standard deviations of its average about 95 per cent of the time. Thus, while actual returns are centred around 8 per cent, they typically vary from a gain of 48 per cent to a loss of 32 per cent.)
- The second asset class is a safe alternative. With a return of 3 per cent and a standard deviation of zero. The safe alternative can be either a bank savings account or a government bond.

FIGURE 25.3

The Trade-Off Between Risk and Return

When people increase the percentage of their savings that they have invested in shares, they increase the average return they can expect to earn, but they also increase the risks they face.



Each point in this figure represents a particular allocation of a portfolio between risky shares and the safe asset. The figure shows that the more a person puts into stock, the greater is both the risk and the return.

Acknowledging the risk–return trade-off does not, by itself, tell us what a person should do. The choice of a particular combination of risk and return depends on a person’s risk aversion, which reflects a person’s own preferences. But it is important for shareholders to realize that the higher average return that they enjoy comes at the price of higher risk.

SELF TEST Describe three ways that a risk-averse person might reduce the risk she faces.

CASE STUDY The Quants and Zero Risk

Having an economics degree may give you an advantage in the employment stakes but many financial institutions have also found highly skilled mathematicians and physicists extremely useful. They became known as ‘the quants’. One such ‘quant’ was David Li, a talented mathematician who worked on quantitative analysis at the Wall Street investment bank J.P. Morgan Chase. In 2000 he published a paper in the *Journal of Fixed Income* entitled ‘On Default Correlation: A Copula Function Approach’.

Li had some background in actuarial science as well as business and in his role within J.P. Morgan Chase, he was part of a growing number of highly qualified mathematical and statistical individuals being employed by financial institutions to identify risk and find ways of reducing it to a minimum – if not eliminating it totally. The basis behind this ‘quantitative finance’ was that the market could be beaten. What quantitative finance provided was the possibility of improved information that enabled decision-makers to invest with significantly reduced risk.

The basis of Li’s paper in 2000 focused on the problems of probabilities and correlation of events. For example, a company in the southwest of England sources animal hide and processes it into a range of different quality leather for sale to manufacturers of leather products. It sources its hides from a company in Kenya. Analysis of data can give reasonably reliable estimates of the probability of both these businesses’ chances of failure. The probability of the UK company failing might be 0.3 and the probability of the Kenyan company failing, 0.4. However, given the relationship between the two, if the Kenyan company did fail then there is a relatively high probability that the UK company might also fail if it cannot source the same high-quality hides elsewhere quickly. The correlation between the two companies failing, therefore, is quite high.

What is the correlation between the chances of the UK company failing and an Icelandic tourism operator also failing? Intuition might suggest that the correlation would be weak given the fact that the two companies are operating in different markets and are not related. However, they may be related by the fact that they have both acquired loans from a bank based in Germany. If that bank suffered problems then the chance of the two companies failing would be relatively high. Li attempted to apply his skills to the effect of the default of one business on another. If this could be modelled then it could help decision-makers to reduce the risk of investing.

One of the statistical devices used for modelling linked events is a copula. The chances of two or more events occurring together can be modelled through the use of a copula which produces a distribution. The specific distribution that Li used was a Gaussian Copula, which to many students is the familiar ‘bell curve’ given by a normal distribution. Li looked at the probability of any two or more elements in a pool of debt defaulting and his formula showed that this was equal to the normal distribution (the copula) times the probabilities of the time taken for the elements to default (referred to as survival times) multiplied by a correlation constant. The survival time aspect of the formula was derived from Li’s knowledge of the work of actuaries in looking at survival rates of partners following the death of one of them. In this scenario events (in this case death) tend to an average and, as mentioned earlier, in this scenario there are certainties – death will occur at some point in the future. With assets the outcomes were not so clear cut and while Li’s formula looked ‘elegant’ it was not able to take into account the range and randomness of

(continued)

events which were not so clear cut. Indeed, Li appeared to recognize (as any good scientist would claim to do) that his formula did have limitations. In 2005, he is quoted as saying ‘Very few people understand the essence of the model’.

For those in the financial world, the model was useful in that it appeared to be able to reduce risk to a ‘simple’ number and thus help price that risk. It told decision-makers what effect the default of one company might have on another regardless of the knowledge of either company. Li’s formula gained some currency in financial markets to the extent that ratings agencies such as Moody’s and Standard & Poor’s incorporated it into their methodology. In particular the formula could be used to assess risk on collateralized debt obligations (CDOs).

Li’s formula may have provided investors with new information which they could use to factor in when making decisions – it became part of the set of ‘relevant generally available information’. The problem was that, despite having this information, those using it may not have fully understood the information which they were exploiting. The whole point of employing top-ranked mathematicians, statisticians and physicists in the world of finance since the latter part of the 1990s was that their skills were very special – it was simply not something that anybody else but these brilliant minds could do; if it had have been possible then others would have done it!

The problem of employing such minds to try and beat the market is that the new information they bring to the market is highly technical, is based on assumptions that are standard in science but which are recognized in that discipline as ‘provisional’, and that any theory or model is simply there to be taken apart and either destroyed or improved upon. The information we now have on the financial crisis, some of the causes of which were laid at the feet of these ‘quants’, is something that markets can now factor into decision-making. There are a great many things in economics that can be modelled and mathematics can be very useful in helping to provide the means to analyse and to predict, but ultimately economics is about human behaviour. Despite such rigorous analysis, the old adage about baskets and eggs still applies!



David Li was referred to by some media outlets as the man who was responsible for the credit crunch and subsequent recession.

ASSET VALUATION

Now that we have developed a basic understanding of the two building blocks of finance – time and risk – let’s apply this knowledge. This section considers a simple question: what determines the price of a share of stock? Like most prices, the answer is supply and demand. But that is not the end of the story. To understand share prices, we need to think more deeply about what determines a person’s willingness to pay for a share of stock.

Fundamental Analysis

Let’s imagine that you have decided to put 60 per cent of your savings into company shares and, to achieve diversification, you have decided to buy 20 different shares. How should you pick the 20 for your portfolio?

When you buy company shares, you are buying shares in a business. When deciding which businesses you want to own, it is natural to consider two things: the value of the business and the price at which the shares are being sold. If the price is less than the value, the stock is said to be *undervalued*. If the price is more than the value, the stock is said to be *overvalued*. If the price and the value are equal, the stock is said to be *fairly valued*. When choosing 20 stocks for your portfolio, you should prefer undervalued stocks. In these cases, you are getting a bargain by paying less than the business is worth.

This is easier said than done. Learning the price is easy: you can just look it up in the newspaper or online or have it downloaded to an app on your mobile device. Determining the value of the business is the hard part. The term **fundamental analysis** refers to the detailed analysis of a company to determine its value. Many financial sector firms hire stock price analysts to conduct such fundamental analysis and offer advice about which stocks to buy.

fundamental analysis the study of a company's accounting statements and future prospects to determine its value

The value of a stock to a shareholder is what he gets out of owning it, which includes the present value of the stream of dividend payments and the final sale price. Recall that *dividends* are the cash payments that a company makes to its shareholders. A company's ability to pay dividends, as well as the value of the stock when the shareholder sells his shares, depends on the company's ability to earn profits. Its profitability, in turn, depends on a large number of factors – the demand for its product, how much competition it faces, how much capital it has in place, whether its workers are unionized, how loyal its customers are, what kinds of government regulations and taxes it faces, and so on. The job of fundamental analysts is to take all these factors into account to determine how much a share of stock in the company is worth.

If you want to rely on fundamental analysis to pick a stock portfolio, there are three ways to do it. One way is to do all the necessary research yourself, by reading through companies' annual reports and so forth. A second way is to rely on the advice of financial analysts. A third way is to buy into an investment fund, which has a manager (sometimes called a fund manager) who conducts fundamental analysis and makes the decision for you.

The Efficient Markets Hypothesis

There is another way to choose 20 shares for your portfolio: pick them randomly by, for instance, putting the stock pages of the *Financial Times* on your notice board and throwing darts at the page. This may sound crazy, but there is reason to believe that it won't lead you too far astray. That reason is called the **efficient markets hypothesis**.

efficient markets hypothesis the theory that asset prices reflect all publicly available information about the value of an asset

To understand this theory, the starting point is to acknowledge that each company listed on a major stock exchange is followed closely by many fund managers, such as the individuals who run investment funds. Every day, these managers monitor news stories and conduct fundamental analysis to try to determine the fund's value. Their job is to buy a share when its price falls below its value, and to sell it when its price rises above its value.

The second piece to the efficient markets hypothesis is that the equilibrium of supply and demand sets the market price. This means that, at the market price, the number of shares being offered for sale exactly equals the number of shares that people want to buy. In other words, at the market price, the number of people who think the share is overvalued exactly balances the number of people who think it's undervalued. As judged by the typical person in the market, all shares are fairly valued all the time.

According to this theory, the stock market is **informationally efficient**: It reflects all available information about the value of the asset. Share prices change when information changes. When the good news about the company's prospects become public, the value and the stock price both rise. When the company's prospects deteriorate, the value and price both fall. But at any moment in time, the market price is the best guess of the company's value based on available information.

informationally efficient reflecting all available information in a rational way

One implication of the efficient markets hypothesis is that stock prices should follow a **random walk**. This means that changes in stock prices are impossible to predict from available information. If, based on publicly available information, a person could predict that a stock price would rise by 10 per cent tomorrow, then the stock market must be failing to incorporate that information today. According to this theory, the only thing that can move stock prices is news that changes the market's perception of the company's value. But news must be unpredictable – otherwise, it wouldn't really be news. For the same reason, changes in stock prices should be unpredictable.

random walk the path of a variable whose changes are impossible to predict

If the efficient markets hypothesis is correct, then there is little point in spending many hours studying the business page to decide which 20 shares to add to your portfolio. If prices reflect all available information, no stock is a better buy than any other. The best you can do is buy a diversified portfolio.

Efficient Markets Hypothesis: Dead and Buried or Just Resting? Any theory is a framework for analysing the world and which allows us to be able to make predictions. Theories are subject to testing and can be found to be inaccurate, limited in understanding or just plain wrong! One of the problems with economic theories, unlike those of the hard sciences, is that they are not easily testable in a laboratory environment where experiments can be repeated in order to establish the reliability or otherwise of a theory. Economic theories are used as the basis for policy making and if a theory is inaccurate or unreliable then policies and decisions can be undermined. The assumption of efficient markets and the belief that markets self-correct and revert to reflect true market value was the basis of the regulatory framework that underpinned the financial system in most major centres of the world.

New information comes out all the time but to the extent that this information really is new, it is unpredictable. In hindsight it always seems that we could have foreseen events after they have arisen; a chain of causality can invariably be established. However, if humans were able to predict future events with any certainty or predict the news it would not be news, by definition!

Much of the above reasoning, however, is based on the principle of rational behaviour. The question raised in the wake of the financial crisis was the extent to which the 'bubbles' which occurred in asset prices represented rational behaviour as opposed to a herd mentality or mass psychology as proposed by the Yale economist and 2013 Nobel Prize winner, Robert Shiller, and Richard Thaler at the University of Chicago, who are leading behavioural economists. The existence of speculative bubbles suggests that markets react to what Alan Greenspan, a former chair of the US Federal Reserve, called 'irrational exuberance' and what Keynes referred to as 'animal spirits'. In speculative bubbles, asset prices rise, not because of the stream of income that flows from them and the price when sold, but solely because of an expectation of what others will think the asset will be worth in the future. It is worth pausing on this for a moment, because there is a subtle difference between valuing an asset based on fundamental analysis – the present value of future dividends – and valuing it based on an assumption that people will bid up the price *regardless of expected future dividends*. It is when the price gets detached from the fundamentals and starts to be contingent only on what everyone in the market expects everyone else will think that a bubble starts to form.

Keynes believed that, given the fact that most investors would sell shares they own at some point in the future, it was not unreasonable to have some concern about others' valuation of that asset. Such views could lead to irrational waves of optimism and pessimism. The efficient markets hypothesis (EMH)

is based on an assumption that there are a sufficient number of people in the market who act rationally to counter the few that do not.

Central to the EMH is the role of information. Information is necessary to value anything, to enable the valuer to make judgements. This information could include data of different kinds. It might include statistical data based on historical records or on reliable forecasting techniques; data can be gained from official statements from a business, from trends, market research, financial audits, analyst reports on the business in relation to its market and the economy, from government and its agencies on economic performance, from other specialist market analysts such as Mintel and Experian, representatives of retail trades and so on. In valuing the price of an asset, therefore, the assumption is that the market is informationally efficient in that it reflects all available information about the value of that asset.

The extent to which this assumption can be accepted is open to question. Access to information and the speed at which information travels is now greater than ever before. Despite the growth in technology, information transfer is not instantaneous nor is it assimilated and understood by all at the same speed and with the same depth. There is, therefore, a time lag involved with information transfer. As we have seen above, if individuals are able to exploit that time lag they can use this to their advantage to make profits; this is the basis of arbitrage. A conclusion from this might be that EMH implies that markets will clear and that most of the time, or on average, that they are efficient. The remaining time periods, when they are not efficient, may be where problems can arise.

A belief in the fundamentals of EMH led even revered financial luminaries such as Alan Greenspan to have to admit that he may have been wrong in some of his assumptions.

In testimony given to the Committee of Government Oversight and Reform in the US on 23 October 2008 in the wake of the financial crisis, Greenspan said:

... those of us who have looked to the self-interest of lending institutions to protect shareholders' equity (myself especially) are in a state of shocked disbelief. Such counterparty surveillance is a central pillar of our financial markets' state of balance. If it fails, as occurred this year, market stability is undermined.

In response to a further question from Committee Chairman, Henry Waxman, who asked:

You found that your view of the world, your ideology, was not right, it was not working?

Greenspan replied:

That's precisely the reason I was shocked because I'd been going for 40 years or so with considerable evidence that it was working exceptionally well.

Information may be widely available and extensive but it does not mean that it is always understood. Decision-makers may have access to a wide range of information but some vital part of that information may be missed if understanding is not complete. Greenspan alluded to this in his testimony. Information may not only be misunderstood but also be partial; gathering information requires some investment in terms of research (information gathering and processing) on the part of decision-makers and this investment can be expensive. In any decision-making, therefore, research costs have to be built in. One very good example of this principle will be considered in the next section. It looks at the way in which information can be made available to decision-makers, can be used and exploited but without critical acceptance may provide the basis for erroneous judgements, as we saw in our outline of the role of 'the quants' in the Case Study earlier in this chapter.

Faith in markets is fundamental to economic growth in most of the world and the financial crisis of 2007–2009 has shaken this faith to the core. There have been many who have consigned EMH to the 'dustbin of history' but there are others that suggest that the events since the early 2000s are testament to the robustness of EMH – that the market cannot be beaten and that attempts to try and do so will ultimately fail. The development of sophisticated information to try and beat the market is one example of how this has happened. The use of such techniques to make 'superrational' valuations of assets is impossible, its supporters would argue.

What may come out of the analysis of the financial crisis is a better understanding of how markets work, especially in a rapidly changing world. The assumptions that have been made about how markets work, the variables that are considered, the effect of external shocks and the extent to which humans may act irrationally, are risk seeking or risk averse and more, may all be revisited, revised and re-interpreted. Ultimately we may be in a better position to understand the complexity of markets but to arrive at the conclusion that our understanding and our mastery is complete is akin to a pronouncement by physicists that they understand every aspect of the physical world and the universe; in short it is not going to happen!

Minsky's Financial Instability Hypothesis There are some, therefore, who suggest that the EMH has been found wanting. Some have suggested an alternative and referred to the idea of American 20th-century economists, Hyman Minsky and Irving Fisher, of the financial instability hypothesis (FIH). This suggests that the idea that markets always move to an equilibrium position and remain in that position pending some external shock may have some resonance in commodities or goods markets but that financial markets are not subject to the same forces. Minsky and Fisher argue that financial markets create their own internal forces which create periods of asset inflation and credit expansion. This will be followed by contraction in credit and asset deflation. Financial markets, they conclude, are not self-optimizing or stable and far from allocating resources efficiently the outcome may be sub-optimal. Minsky suggests that some of these internal forces relate to the lack of supply of assets which drive demand in those markets. For example, the lack of supply of housing drives demand for housing and forces prices upwards creating a bubble. Changing asset prices (such as houses) in turn act as a driver for demand for those assets. If house prices are rising quickly there is an incentive for buyers to want to get into the market quickly to avoid having to pay higher prices and also to benefit from rising prices once they have purchased. This simply fuels demand further in the face of limited supply and so drives price up higher.

Maybe the last word on this issue can be with Burton G. Malkiel from Princeton who, in a paper in 2003, concluded:

As long as stock markets exist, the collective judgment of investors will sometimes make mistakes. Undoubtedly, some market participants are demonstrably less than rational. As a result, pricing irregularities and predictable patterns in stock returns can appear over time and even persist for short periods. Moreover, the market cannot be perfectly efficient or there would be no incentive for professionals to uncover the information that gets so quickly reflected in market prices, a point stressed by Grossman and Stiglitz (1980). Undoubtedly, with the passage of time and with the increasing sophistication of our databases and empirical techniques, we will document further apparent departures from efficiency and further patterns in the development of stock returns.

But I suspect that the end result will not be an abandonment of the belief of many in the profession that the stock market is remarkably efficient in its utilization of information. Periods such as 1999 where 'bubbles' seem to have existed, at least in certain sectors of the market, are fortunately the exception rather than the rule. Moreover, whatever patterns or irrationalities in the pricing of individual stocks that have been discovered in a search of historical experience are unlikely to persist and will not provide investors with a method to obtain extraordinary returns. If any \$100 bills are lying around the stock exchanges of the world, they will not be there for long.

Malkiel, B.G. (2003). 'The Efficient Market Hypothesis and its Critics'. CEPS Working Paper no. 91.

Market Irrationality

The efficient markets hypothesis assumes that people buying and selling stock rationally process the information they have about the stocks' underlying value. But is the stock market really that rational? Or do share prices sometimes deviate from reasonable expectations of their true value?

There is a long tradition suggesting that fluctuations in share prices are partly psychological. In the 1930s, British economist John Maynard Keynes suggested that asset markets are driven by the 'animal spirits' of investors – irrational waves of optimism and pessimism. Alan Greenspan's reference to 'irrational exuberance' became popular in describing stock market behaviour, especially when it was used as the title of a book on stock market behaviour by Robert Shiller that successfully predicted a large fall in US stock prices. Nevertheless, whether the exuberance of stock markets in the USA and around the world in the 1990s and the middle of the noughties was irrational given the information available at the time remains debatable.

Sustained and sometimes rapid rises in stock prices are called speculative bubbles, since often the price rise comes to an end and share prices fall abruptly, just as a soap bubble will rise and rise until it suddenly bursts. The possibility of such speculative bubbles arises in part because the value of a share to a stockholder depends not only on the stream of dividend payments but also on the final sale price of the share. Thus, a person might be willing to pay more than a stock is worth today if he expects another

person to pay even more for it tomorrow. When you evaluate a stock, you have to estimate not only the value of the business but also what other people will think the business is worth in the future.

There is much debate among economists about whether departures from rational pricing are important or rare. Believers in market irrationality point out (correctly) that the stock market often moves in ways that are hard to explain on the basis of news that might alter a rational valuation. We have already described a number of examples where disciplines such as biology and psychology are helping to contribute to our understanding of human behaviour. The growth in 'neuroeconomics' is one example of where such information is helping to push the boundaries of our understanding and is being applied to finance. Believers in the efficient markets hypothesis point out (correctly) that it is impossible to know the correct, rational valuation of a company, so one should not quickly jump to the conclusion that any particular valuation is irrational. Moreover, if the market were irrational, a rational person should be able to take advantage of this fact; yet, as we have seen, beating the market is nearly impossible.

SELF TEST Investors sometimes refer to 'highly respected' companies as 'blue chip companies'. These are usually large, national or multinational corporations with a reputation for high-quality management or products. Examples of blue chip companies quoted on the London Stock Exchange would be Rolls-Royce, Unilever and BP. According to the efficient markets hypothesis, if you restrict your stock portfolio to blue chip companies, would you earn a better than average return? Explain.

FYI



Keynes's View of Stock Markets

As well as being the most influential economist of the 20th century, Keynes was also a successful stock market investor and considerably enriched his college in Cambridge by managing its investment portfolio. Yet his opinion of professional investment managers was notoriously low. In his *Treatise on Money* (1930), he wrote:

... the vast majority of those who are concerned with the buying and selling of securities know almost nothing whatever about what they are doing. They do not possess even the rudiments of what is required for a valid judgment, and are the prey of hopes and fears easily aroused by transient events and as easily dispelled. This is one of the odd characteristics of the capitalist system under which we live, which, when we are dealing with the real world, is not to be overlooked.

Keynes was not known for his modesty, but, even so, one suspects that he may have been overstating his case somewhat for the sake of emphasis. In *The General Theory of Employment, Interest and Money* (1936), he sketched out a more balanced view that amounts to a simple theory of speculative bubbles in a very famous passage in which he compares the stock market to a beauty contest:

... professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not

those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be.

For these reasons, a person who believes that departures from rational pricing are the rule rather than the exception in the stock market are sometimes referred to as having a Keynesian view of financial markets.

CONCLUSION

This chapter has developed some of the basic tools that people should (and often do) use as they make financial decisions. The concept of present value reminds us that a euro in the future is less valuable than a euro today, and it gives us a way to compare sums of money at different points in time. The theory of risk management reminds us that the future is uncertain and that risk-averse people can take precautions to guard against this uncertainty. The study of asset valuation tells us that the stock price of any company should reflect its expected future profitability.

Although most of the tools of finance are well established, there is more controversy about the validity of the efficient markets hypothesis and whether stock prices are, in practice, rational estimates of a company's true worth. Rational or not, the large movements in stock prices that we observe have important macroeconomic implications. Stock market fluctuations often go hand in hand with fluctuations in the economy more broadly. We will look at the stock market again when we study economic fluctuations later in the book.

IN THE NEWS



The Efficient Markets Hypothesis

The financial crisis of 2007–2009 led to a number of economists and commentators being fiercely critical of the EMH. Amidst this criticism, what did the founder of the EMH think? Eugene Fama, a professor of finance at the University of Chicago Booth School of Business and joint-winner of the Nobel Prize for Economists in 2013, developed portfolio theory and is credited with developing the theory that markets are efficient.

EMH in Hindsight

Since the financial crisis it has been fashionable to criticize the efficient markets hypothesis as being incorrect. The founder of the EMH, Eugene Fama, has been asked about whether his view of the EMH has changed in the wake of the crisis. In various interviews he has offered a robust and objective analysis of the crisis and what it means for the theory he developed. It is worthwhile clarifying a point that is sometimes missed – in the period leading up to the crisis, people may have assumed that prices being quoted on stock markets reflected the fundamental value of corporations. Quants were employed by finance houses across the world to find ways to beat the market and the development of highly sophisticated statistical and

mathematical models were supposed to help exploit informational inefficiency in markets to give dealers an edge over their rivals. Perhaps one thing that has not been made clear in the wake of the crisis is that in market trades there will always be a winner and a loser. If one trader manages to exploit undervaluation, someone on the opposite side of the trade must have been overvaluing that same trade in order for the exchange to take place.

Hindsight is always an impressive thing – we look back on price rises in the housing market and in asset prices around 2004–2007 and can confidently state that this was an example of an asset and credit bubble. Fama recognizes the importance of hindsight. If investors are living through a period of a bubble and can identify this *at the time*, then this will be part

of known information and so investors would respond in kind. During a credit bubble it makes sense to sell when prices are high in the expectation and knowledge that prices are going to fall. So, if a credit bubble did exist at the time, someone, somewhere would have made a great deal of money. Fama argues that there are always instances where people are going to say that asset prices are too high and that the markets are experiencing a bubble. Equally, there are those who present an opposing argument and both cannot be correct. If prices do collapse then it is relatively easy to pick out the individuals who warned of a bubble and laud them whilst we castigate those who provided an opposing point of view. At the time of writing (February 2013), stock prices have been rising on world markets – are stock markets experiencing a bubble?

A search of the financial media at this time suggested that there are those who believe in a bubble and those who scoff at the very idea. By the time this book is published we may know who was right and who was wrong – or indeed if both were incorrect. Fama’s view is that the word ‘bubble’ has little meaning and that observations on whether markets are experiencing a bubble are right and wrong in equal measures.

Observers might question this view and note that there was never any rationality to lending money to people with very poor credit ratings. Fama counters that part of the sub-prime problems was down to US government policy to promote home ownership among all classes of society and that it was not just the collapse in sub-prime house prices that caused the financial crisis, but a global fall in house prices which took place across most house price bands, not simply those linked to sub-prime. The consequences of the decline in global house prices were problems for the banking sector and a lack of availability of credit in markets. Fama points to the cause being an economic one – a global downturn in economic activity which led to people in all walks of life being unable to meet their credit (and mortgage) commitments. There is always going to be a group of people who are at the margin – just able to afford their commitments provided things don’t change – but when they do these people at the margin are the first to default; they are literally living on the edge. Fama’s view is that the downturn in economic activity started to show itself before the financial crisis and that there was a worldwide fall in house prices that pre-dated the financial crisis. What actually causes recessions is something Fama thinks is still very much

a mystery to macroeconomists and that it can be argued that the financial markets were a victim of the recession and not the cause of it.

Fama is still a passionate believer in the role of financial markets providing a conduit from those who wish to save to those who wish to borrow – including corporates – and that financial innovation of the type that generated CDOs and CDS have contributed much to the development of countries across the globe and to the welfare of millions of people over the last 30 years. If the EMH is about valuing stock prices using all available information, there is an assumption that markets will understand completely the models which they are working with and the information which they have available to price assets. It is highly unlikely that every decision-maker in financial markets knows everything about what they are working with and so there is always some element

of unknown information. Those fund managers working with David Li’s formula may not have fully understood the model but chose to use it anyway, and, as such, decisions are being made with less than complete information.

So, when financial markets get into trouble Fama does not believe that the finger should be solely pointed at the markets themselves or any underlying theory of market efficiency, but he does think that the idea that financial institutions can be ‘too big to fail’ is of real concern because it creates moral hazard and adverse selection. Financial institutions should be allowed to fail and that would be an ultimate regulation.

So, despite the criticism that has been levelled against the EMG, Fama still believes that the fundamental conclusion of the theory, you cannot beat the market, still holds and that financial markets can be described as ‘efficient’. Fama does

Eugene Fama was acknowledged for his work on the empirical analysis of asset prices when he shared the Nobel Prize for Economic Sciences in 2013, along with Robert Shiller and Lars Peter Hansen.



think that the developments in other parts of the University of Chicago on behavioural explanations is an interesting development and, if empirical evidence can be put forward, has much to contribute to the efficient markets debate.

Questions

1 One of the key conclusions of the EMH is that 'you can't beat the market'. Is the fact that financial institutions employed highly qualified maths, physics and

statistics graduates to develop ways to exploit information asymmetries testament to the view that few working in financial markets believed in the EMH? Explain your reasoning.

2 Must there always be a winner and a loser in financial transactions? Explain.

3 Do you think that the word 'bubble' is only ever useful in hindsight? Explain.

4 'Lending to the sub-prime market must have been a rational

thing to do at the time otherwise people would not have done it'. Comment on this view.

5 Fama argues that economists have been researching the cause of the Great Depression in the 1930s for many years but have not yet managed to come up with a definitive conclusion about its causes. Can economists ever identify cause and effect when analysing financial crises such as that in 2007–2009?

SUMMARY

- Because savings can earn interest, a sum of money today is more valuable than the same sum of money in the future. A person can compare sums from different times using the concept of present value. The present value of any future sum is the amount that would be needed today, given prevailing interest rates, to produce that future sum.
- Because of diminishing marginal utility, most people are risk averse. Risk-averse people can reduce risk using insurance, through diversification, and by choosing a portfolio with lower risk and lower return.
- The value of an asset, such as a share of stock, equals the present value of the cash flows the owner of the share will receive, including the stream of dividends and the final sale price. According to the efficient markets hypothesis, financial markets process available information rationally, so a stock price always equals the best estimate of the value of the underlying business. Some economists question the efficient markets hypothesis, however, and believe that irrational psychological factors also influence asset prices.

QUESTIONS FOR REVIEW

- 1 Outline the concept of present value.
- 2 The interest rate is 7 per cent. Use the concept of present value to compare €200 to be received in ten years and €300 to be received in 20 years.
- 3 What benefit do people get from the market for insurance? What two problems impede the insurance company from working perfectly?
- 4 What is meant by the term 'fundamental analysis'?
- 5 What is diversification? Does a shareholder get more diversification going from 1 to 10 stocks or going from 100 to 120 stocks?
- 6 Comparing company shares and government bonds, which has more risk? Which pays a higher average return?
- 7 What factors should a stock analyst think about in determining the value of a share of stock?
- 8 Describe the efficient markets hypothesis, and give a piece of evidence consistent with this theory.
- 9 Explain the view of those economists who are sceptical of the efficient markets hypothesis.
- 10 Why is knowledge of whether people are risk averse or risk seeking important in financial markets?

PROBLEMS AND APPLICATIONS

- 1 About 400 years ago, Native Americans sold the island of Manhattan for \$24. If they had invested this money at an interest rate of 7 per cent per year, how much would they have today?
- 2 A company has an investment project that would cost €10 million today and yield a payoff of €15 million in four years.
 - a. Should the firm undertake the project if the interest rate is 11 per cent? 10 per cent? 9 per cent? 8 per cent?
 - b. Can you figure out the exact cut-off for the interest rate between profitability and non-profitability?
- 3 For each of the following kinds of insurance, give an example of behaviour that can be called *moral hazard* and another example of behaviour that can be called *adverse selection*:
 - a. health insurance
 - b. car insurance.
- 4 Imagine that you intend to buy a portfolio of ten shares with some of your savings. Should the shares be of companies in the same industry? Should the shares be of companies located in the same country? Explain.
- 5 For which kind of stock would you expect to pay the higher average return: stock in an industry that is very sensitive to economic conditions (such as a car manufacturer) or stock in an industry that is relatively insensitive to economic conditions (such as a water company)? Why?
- 6 A company faces two kinds of risk. An idiosyncratic risk is that a competitor might enter its market and take some of its customers. An aggregate risk is that the economy might enter a recession, reducing sales. Which of these two risks would more likely cause the company's shareholders to demand a higher return? Why?
- 7 You have two flatmates who invest in the stock market.
 - a. One flatmate says she buys stock only in companies that everyone believes will experience big increases in profits in the future. How do you suppose the price – earnings ratio of these companies compares to the price – earnings ratio of other companies? What might be the disadvantage of buying stock in these companies?
 - b. Another flatmate says she only buys stock in companies that are cheap, which she measures by a low price-earnings ratio. How do you suppose the earnings prospects of these companies compare to those of other companies? What might be the disadvantage of buying stock in these companies?
- 8 When company executives buy and sell stock based on private information they obtain as part of their jobs, they are engaged in *insider trading*.
 - a. Give an example of inside information that might be useful for buying or selling stock.
 - b. Those who trade shares based on inside information usually earn very high rates of return. Does this fact violate the efficient market hypothesis?
 - c. Insider trading is illegal. Why do you suppose that is?
- 9 Financial markets in the early years of the 21st century were heavily influenced by the work of quantitative analysts ('quants') who looked at ways to reassess the pricing of risk. Discuss some of the advantages and disadvantages of pricing risk using mathematics and mathematical models.
- 10 How might you go about pricing the risk on the following:
 - a. the price of a ten-year bond from the Portuguese government
 - b. the price of a ten-year bond from the German government
 - c. the construction of a new oil well off the South African coast.

26

THE MONETARY SYSTEM

When you walk into a restaurant to buy a meal, you receive something of value: the enjoyment of the food together with the experience of eating the meal in a pleasant environment and being waited upon. To pay for this service, you might hand the restaurateur several pieces of brightly coloured paper decorated with strange symbols, buildings and maps of Europe (or even a portrait of a monarch). Or you might hand him a small, rectangular piece of coloured plastic with a magnetic strip and a small computer chip set into it, which the restaurateur will later return to you. Whether you pay by cash, debit card or credit card, the restaurateur is happy to work hard to satisfy your gastronomic desires either in exchange for pieces of paper that, in and of themselves, are worthless, or in exchange for borrowing an equally worthless small piece of plastic from you for a few minutes.

To anyone who has lived in a modern economy, this social custom is not at all odd. Even though paper money has no intrinsic value, the restaurateur is confident that, in the future, some third person will accept it in exchange for something that the restaurateur does value. And that third person is confident that some fourth person will accept the money, with the knowledge that yet a fifth person will accept the money ... and so on. To the restaurateur and to other people in our society, your cash represents a claim to goods and services in the future. If you paid by debit card, the restaurateur is happy because he knows that money has been transferred more or less instantly to his bank account from your bank account, and the figures on his bank balance also represent a claim to goods and services produced by the economy. If you paid with a credit card, he knows that his account has been credited with the money that you have borrowed on the credit card in order to pay the restaurant bill.

The social custom of using money for transactions is extraordinarily useful in a large, complex society. Imagine, for a moment, that there was no item in the economy widely accepted in exchange for goods and services. People would have to rely on **barter** – the exchange of one good or service for another – to obtain the things they need. To get your restaurant meal, for instance, you would have to offer the restaurateur something of immediate value. You could offer to wash some dishes, clean his car or give him the secret recipe for your family's favourite dish. An economy that relies on barter will have trouble allocating its scarce resources efficiently. In such an economy, trade is said to require the **double coincidence of wants** – the unlikely occurrence that two people each have a good or service that the other wants.

barter the exchange of one good or service for another

double coincidence of wants a situation in exchange where two people each have a good or service that the other wants and can thus enter into an exchange

The existence of money makes trade easier. The restaurateur does not care whether you can produce a valuable good or service for him. He is happy to accept your money, knowing that other people will do the same for him. Such a convention allows trade to be roundabout. The restaurateur accepts your money and uses it to pay his chef; the chef uses her salary to send her child to the crèche; the crèche

uses this money to pay a teacher; and the teacher hires you to child-mind when he goes to the cinema on a Saturday evening. As money flows from person to person in the economy, it facilitates production and trade, thereby allowing each person to specialize in what he or she does best and raising everyone's standard of living.

In this chapter we begin to examine the role of money in the economy. We discuss what money is, the various forms that money takes, how the banking system helps create money, and how the central bank controls the quantity of money in circulation. Because money is so important in the economy, we devote much effort in the rest of this book to learning how changes in the quantity of money affect various economic variables, including inflation, interest rates, production and employment. Consistent with our long-run focus in the previous three chapters, in the next chapter we will examine the long-run effects of changes in the quantity of money. The short-run effects of monetary changes are a more complex topic, which we will take up later in the book. This chapter provides the background for all of this further analysis.

THE MEANING OF MONEY

Economists use the word in a specific sense: **money** is the set of assets in the economy that people regularly use to buy goods and services from other people. The cash in your wallet is money because you can use it to buy a meal at a restaurant or a shirt at a clothes shop. By contrast, if you happened to own part of Facebook, as Mark Zuckerberg does, you would be wealthy, but this asset is not considered a form of money. You could not buy a meal or a shirt with this wealth without first obtaining some cash. According to the economist's definition, money includes only those few types of wealth that are regularly accepted by sellers in exchange for goods and services.

money the set of assets in an economy that people regularly use to buy goods and services from other people

The Functions of Money

Money has three functions in the economy: it is a medium of exchange; a unit of account; and a store of value. These three functions together distinguish money from other assets in the economy, such as stocks, bonds, residential property and art. Let's examine each of these functions of money in turn.

A Medium of Exchange A **medium of exchange** is an item that buyers give to sellers when they purchase goods and services. When you buy a shirt at a clothes shop, the shop gives you the shirt and you give the shop your money. This transfer of money from buyer to seller allows the transaction to take place. When you walk into a shop, you are confident that the shop will accept your money for the items it is selling because money is the commonly accepted medium of exchange.

medium of exchange an item that buyers give to sellers when they want to purchase goods and services

A Unit of Account A **unit of account** is the yardstick people use to post prices and record debts. When you go shopping, you might observe that a shirt is priced at €20 and a ham-and-cheese sandwich at €2. Even though it would be accurate to say that the price of a shirt is 10 sandwiches and the price of a sandwich is 1/10 of a shirt, prices are never quoted in this way. Similarly, if you take out a loan in euros from a bank, the size of your future loan repayments will be measured in euros, not in a quantity

of goods and services. When we want to measure and record economic value, we use money as the unit of account.

unit of account the yardstick people use to post prices and record debts

A Store of Value A **store of value** is an item that people can use to transfer purchasing power from the present to the future. When a seller accepts money today in exchange for a good or service, that seller can hold the money and become a buyer of another good or service at another time. Of course, money is not the only store of value in the economy, for a person can also transfer purchasing power from the present to the future by holding other assets. The term **wealth** is used to refer to the total of all stores of value, including both money and non-monetary assets.

store of value an item that people can use to transfer purchasing power from the present to the future
wealth the total of all stores of value, including both money and non-monetary assets

Liquidity

Economists use the term **liquidity** to describe the ease with which an asset can be converted into the economy's medium of exchange. Because money is the economy's medium of exchange, it is the most liquid asset available. Other assets vary widely in their liquidity. Most stocks and bonds can be sold easily with small cost, so they are relatively liquid assets. By contrast, selling a car or a Rembrandt painting requires more time and effort, so these assets are less liquid.

liquidity the ease with which an asset can be converted into the economy's medium of exchange

When people decide in what form to hold their wealth, they have to balance the liquidity of each possible asset against the asset's usefulness as a store of value. Money is the most liquid asset, but it is far from perfect as a store of value. When prices rise, the value of money falls. In other words, when goods and services become more expensive, each euro or pound in your pocket can buy less. This link between the price level and the value of money will turn out to be important for understanding how money affects the economy.

The Kinds of Money

When money takes the form of a commodity with intrinsic value, it is called **commodity money**. The term *intrinsic value* means that the item would have value even if it were not used as money. One example of commodity money is gold. Gold has intrinsic value because it is used in industry and in the making of jewellery. Although today we no longer use gold as money, historically gold has been a common form of money because it is relatively easy to carry, measure, and verify for impurities. When an economy uses gold as money (or uses paper money that is convertible into gold on demand), it is said to be operating under a *gold standard*.

commodity money money that takes the form of a commodity with intrinsic value

Although gold has, historically, been the most common form of commodity money, other commodity monies have been used from time to time. For example, in the hyperinflation in Zimbabwe in the early 2000s the country's people began to lose faith in the Zimbabwean dollar; people began to trade goods and services with one another using cigarettes as the store of value, unit of account and medium of exchange.

Money without intrinsic value is called **fiat money**. A fiat is simply an order or decree, and fiat money is established as money by government decree. For example paper euros are able to circulate as legal tender in 18 European countries because the governments of those countries have decreed that the euro be valid currency in each of their economies.

fiat money money without intrinsic value that is used as money because of government decree

Although governments are central to establishing and regulating a system of fiat money (by prosecuting counterfeiters, for example), other factors are also required for the success of such a monetary system. To a large extent, the acceptance of fiat money depends as much on expectations and social convention as on government decree. Zimbabweans preferred to accept cigarettes (or American dollars) in exchange for goods and services, because they were more confident that these alternative monies would be accepted by others in the future.

Money in the Economy

As we will see, the quantity of money circulating in the economy, called the **money stock**, has a powerful influence on many economic variables. But before we consider why that is true, we need to ask a preliminary question: what is the quantity of money? In particular, suppose you were given the task of measuring how much money there is in the economy of a country like the United Kingdom, or in the economies of the 18 countries that make up the euro area. What would you include in your measure?

money stock the quantity of money circulating in the economy

The most obvious asset to include is **currency** – the paper notes and metal coins in the hands of the public. Currency is clearly the most widely accepted medium of exchange in a modern economy. There is no doubt that it is part of the money stock.

currency the paper banknotes and coins in the hands of the public

Yet currency is not the only asset that you can use to buy goods and services. Most businesses also accept payment by debit card, which allows money to be transferred electronically between your current account and the current account of the business. Another, more old-fashioned way of transferring money between current accounts is to write a personal cheque, and personal cheques are still accepted as a means of payment although being phased out in many countries.

So is a debit card or a cheque money? Not really – it is the bank account on which the cheque or debit card draws which contains the money. A debit card is just a *means* of transferring money between accounts. The same is true of a cheque because it is simply a *means* of transferring money from one bank account to another.

What about credit cards? We need to think about these even more carefully, because credit cards are not really a method of payment but a method of *deferring* payment. When you buy a meal with a credit card, the bank that issued the card pays the restaurant what it is due – you have effectively borrowed from the bank. At a later date, you will have to repay the bank (perhaps with interest). When the time comes to pay your credit card bill, you will probably do so by direct transfer from your current account. The balance in this current account is part of the economy's stock of money. Notice that credit cards are very different from debit cards, which automatically withdraw funds from a bank account to pay for items bought. Why does the restaurateur accept payment by credit card? Because he gets his money immediately by having his bank account credited for the price of the meal even though you do not have to pay the credit card company back immediately. Again, however, it is the underlying movement in the restaurateur's bank balance that matters.

Thus, although a debit card and a credit card can each be used to settle the restaurant bill, none of them are money – they are each a method of transferring money between bank accounts. In the case of a debit card or a cheque, money is transferred from your account to the restaurateur's account more or less immediately or with a very short lag. In the case of a credit card, the restaurateur gets his money in his account more or less immediately and you will then have to settle up with the bank issuing the credit card later by drawing on your bank account. In every case, the true movement in money occurs when bank balances change.

Wealth held in your current account is almost as convenient for buying things as wealth held in your wallet. To measure the money stock, therefore, you might want to include **demand deposits** – balances in bank accounts that depositors can access *on demand* simply by using their debit card. Once you start to consider balances in current accounts as part of the money stock, you are led to consider the large variety of other accounts that people hold at banks and other financial institutions. Bank depositors usually cannot use their debit cards against the balances in their savings accounts, but they can (mostly) easily transfer funds from savings into current accounts. In addition, depositors in money market funds can often use debit cards against their balances. Thus, these other accounts should plausibly be counted as part of the money stock.

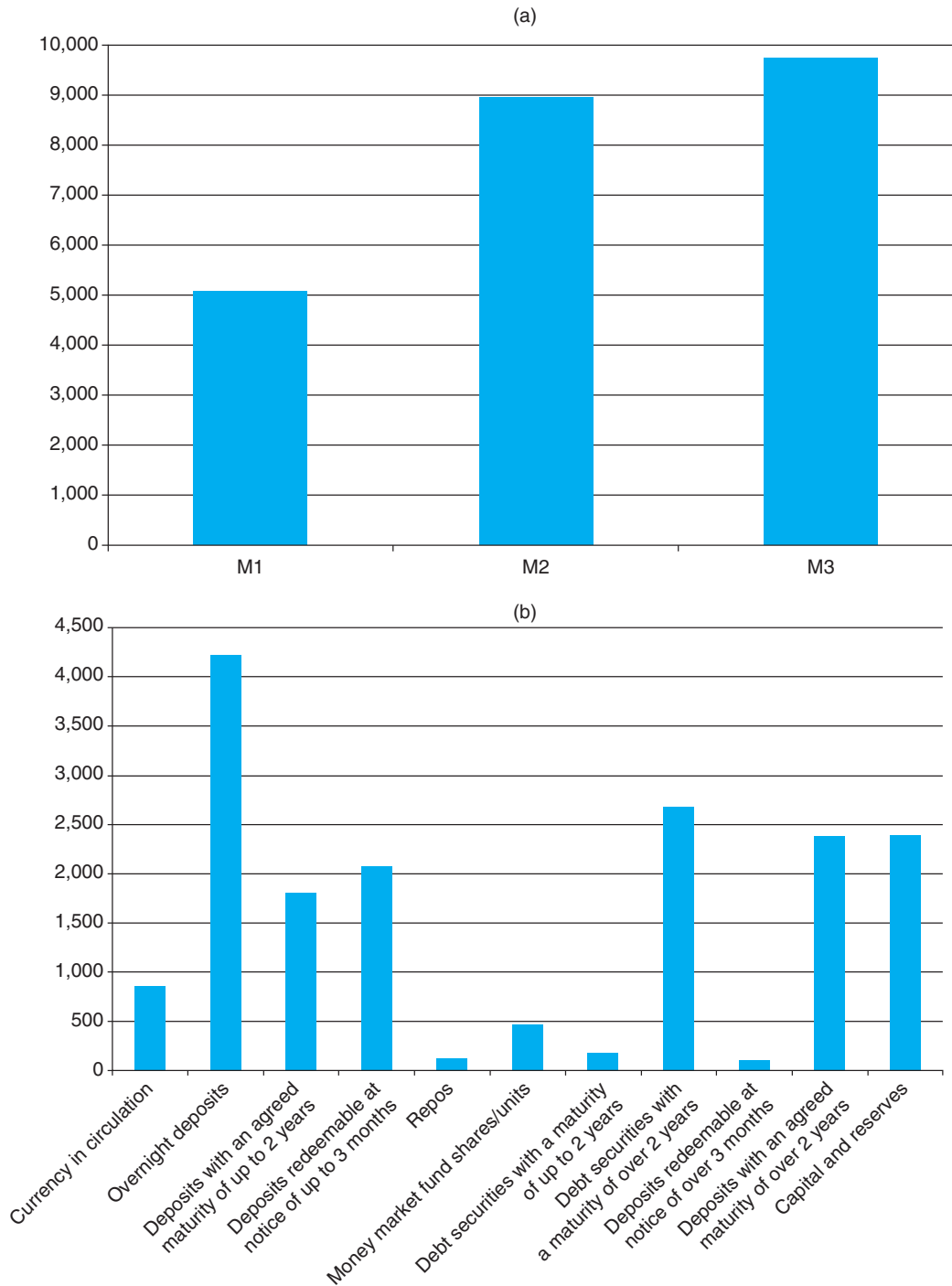
demand deposits balances in bank accounts that depositors can access on demand by using a debit card or writing a cheque

In a complex economy, it is in general not easy to draw a line between assets that can be called 'money' and assets that cannot. The coins in your pocket are clearly part of the money stock, and your car clearly is not, but there are many assets for which the choice is less clear. Therefore, various measures of the money stock are available for advanced economies. Panel (a) of Figure 26.1 shows the three most important measures for the euro area, designated M1 (a 'narrow' measure), M2 (an 'intermediate' measure) and M3 (a 'broad' measure). Each of these measures uses a slightly different criterion for distinguishing between monetary and non-monetary assets. In the UK, the most widely observed measures of the money stock are M0 (a narrower measure than the European M1 and corresponding to notes and coins in circulation plus bankers' balances held with the Bank of England) and M4 (a broad measure, similar – but not identical to – the European M3). Panel (b) of Figure 26.1 shows a breakdown of the various components of the money stock.

For our purposes in this book, we need not dwell on the differences between the various measures of money. The important point is that the money stock for an advanced economy includes not just currency but also deposits in banks and other financial institutions that can be readily accessed and used to buy goods and services.

FIGURE 26.1**Three Measures of the Money Stock for the Euro Area (billions of euro)**

Panel (a) shows the three measures of euro area money stock are M1 (a 'narrow' monetary aggregate that comprises currency in circulation and overnight deposits), M2 (an 'intermediate' monetary aggregate that comprises M1 plus deposits with an agreed maturity of up to two years and deposits redeemable at notice of up to three months) and M3 (a 'broad' monetary aggregate that comprises M2 plus repurchase agreements, money market fund shares and units as well as debt securities with a maturity of up to two years). Panel (b) shows the size of the different components of the money stock. This figure shows the size of each measure in December 2012.



SELF TEST List and describe the three functions of money.

CASE STUDY The 'True' Austrian Money Supply

As with many aspects of economics, definitions of the money supply are open to different interpretations. The role of the money supply is going to be central too much of the analysis in the rest of this book as noted but it is clear that if the measure of the money supply being used varies then it is likely that the analysis and the conclusions that result will also vary. Fortunately, economists of almost every persuasion are in agreement that currency – notes and coins are a part of the money supply. What else should count as money and thus be part of the accepted definition is subject to different interpretations. One example is the so-called 'True' Austrian definition of the money supply put forward by the Austrian economist Ludwig von Mises in his *Theory of Money and Credit* published in 1953. von Mises defined the money supply as standard money + money substitutes. Standard money he defined as notes and coins in circulation or currency.

Money substitutes are more challenging to define but von Mises suggested that money substitutes must be perfectly secure and immediately convertible. In other words, they must be evidence of the ownership of standard money and be a claim to money in the present. By this definition, a number of the elements of the measured money supply by central banks across the world, including the ECB and the Bank of England, would not be considered as being part of the money supply and, according to economists from this school, represent a flawed measure of the money supply.

The reason is that some of these elements represent the individual giving up a claim to money in the present for some promise of money in the future. In order for the claim to be liquidated a further transaction has to be made sometime in the future and as such the credit claim is not immediately convertible. For example, savings deposits at commercial banks would be included as part of M2 in many countries but such deposits have legal terms under which they can be liquidated, for example the provision of a 30-day notice period given to the bank, and as such they represent a credit claim and not a money substitute. Under the Austrian school definition, time deposits and money funds would also be classed as credit claims and thus not money substitutes.

Ludwig von Mises, founder of the so-called Austrian school of economics proposes a different view of the money supply which would impact on policy decisions. An example of where economists disagree as outlined in Chapter 2.



An economist at the Mises Institute in the United States, Dr Frank Shostak, provides some clarification of the difference between a money substitute and a credit claim in an essay published in 2000 (Shostak, F. (2000) *The Mystery of the Money Supply*):

Once an individual places his money in a bank's warehouse he is in fact engaging in a claim transaction. In depositing his money, he never relinquishes his ownership. No one else is expected to make use of it. When Joe stores his money with a bank, he continues to have an unlimited claim against it and is entitled to take charge of it at any time. Consequently, these deposits, labelled demand deposits, are part of money ...

This must be contrasted with a credit transaction, in which the lender of money relinquishes his claim over the money for the duration of the loan. Credit always involves a creditor's purchase of a future good in exchange for a present good. As a result, in a credit transaction, money is transferred from a lender to a borrower.

The debate over what is and what is not money is not simply academic, it is important for policy making; after all, we would not use an unaccepted measure of temperature in situations where maintaining specific level of heat or cold are required because of the disastrous consequences that could result so if we use incorrect measures of the money supply it follows that the consequences could be equally serious.

THE ROLE OF CENTRAL BANKS

Whenever an economy relies on a system of fiat money – as all modern advanced economies do – some agency must be responsible for regulating the system. This agency is generally known as the **central bank** – an institution designed to regulate the quantity of money made available in the economy, called the **money supply**. Two of the most important central banks in Europe are the European Central Bank and the Bank of England. Other major central banks around the world include the US central bank – the Federal Reserve – the Bank of Japan and the People’s Bank of China. We’ll take a closer look at the European Central Bank and the Bank of England in a moment. Before then, however, we can look at some features of central banks in general.

central bank an institution designed to regulate the quantity of money in the economy
money supply the quantity of money available in the economy

The central bank of an economy has the power to increase or decrease the amount of currency in that economy. The set of actions taken by the central bank in order to affect the money supply is known as **monetary policy**. In simple metaphorical terms, you can imagine the central bank printing banknotes and dropping them by helicopter. Similarly, you can imagine the central bank using a giant vacuum cleaner to suck banknotes out of people’s wallets. Although in practice the central bank’s methods for changing the money supply are more complex and subtle than this, the helicopter-vacuum metaphor is a good first approximation to the meaning of monetary policy.

monetary policy the set of actions taken by the central bank in order to affect the money supply

We discuss later in this chapter how the central bank actually changes the money supply, but it is worth noting here that an important tool the central bank can use is **open-market operations** – the purchase and sale of non-monetary assets from and to the banking sector. For example, if the central bank decides to increase the money supply, it can do this by creating currency and using it to buy bonds from the public in the bond market. After the purchase, the extra currency is in the hands of the public. Thus, an open-market purchase of bonds by the central bank increases the money supply. Conversely, if the central bank decides to decrease the money supply, it can do this by selling bonds from its portfolio to the public. After the sale, the currency it receives for the bonds is out of the hands of the public. Thus, an open-market sale of bonds by the central bank decreases the money supply.

open-market operations the purchase and sale of non-monetary assets from and to the banking sector by the central bank

The central bank of an economy is an important institution because changes in the money supply can profoundly affect the economy. One of the *Ten Principles of Economics* is that prices rise when too much money is printed. Another of the *Ten Principles of Economics* is that society faces a short-run trade-off between inflation and unemployment. The power of the central bank rests on these principles. For reasons we discuss more fully in the coming chapters, the central bank’s policy decisions have an important influence on the economy’s rate of inflation in the long run and the economy’s employment and production in the short run. In particular, because of the link between the amount of money in the economy and the inflation rate (the rate of increase of prices), the central bank is often seen as the guardian of price stability in a modern economy charged with the duty to maintain inflation at or near an inflation target – a policy we’ll discuss more fully in the coming chapters. To be precise, the central bank should perhaps be thought

of as the guardian of inflation stability, rather than price stability, since even with a constant, low rate of inflation, prices are by definition rising. Still, if inflation is low and stable, prices might be said to be rising in a stable fashion – and, in any case, this usage is now well established, so we'll follow suit.

THE EUROPEAN CENTRAL BANK AND THE EUROSYSTEM

The **European Central Bank (ECB)**, located in Frankfurt, Germany, was officially created on 1 June 1998 as a number of European countries had decided that they wished to enter European Monetary Union (EMU) and have the same currency – the euro – circulating among them. We'll discuss the pros and cons of monetary union in a later chapter. For now, though, we just note that if a group of countries has the same currency, then it makes sense for the countries in the group to have a common monetary policy, and the ECB was set up for precisely this purpose. There were originally 11 countries making up the euro area: Austria, Belgium, Germany, Spain, France, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Finland. By 2010, there were 16 countries, the new additions being Greece, Cyprus, Slovakia, Slovenia and Malta, Estonia adopted the euro in 2011 and Latvia adopted the euro in January 2014 to make the euro area currently consist of 18 countries.

European Central Bank (ECB) the overall central bank of the 18 countries comprising the European Monetary Union

The primary objective of the ECB is to promote price stability throughout the euro area and to design and implement monetary policy that is consistent with this objective. The ECB operates with the assistance of the national central banks in each of the euro area countries, such as the Banque de France, the Banca d'Italia, the Bank of Greece and the German Bundesbank. The network made up of the ECB together with the 18 euro area national central banks is termed the **Eurosistem**.

Eurosistem the system made up of the ECB plus the national central banks of each of the 18 countries comprising the European Monetary Union

The implementation of monetary policy by the ECB is under the control of the Executive Board, which comprises the President and Vice-President of the ECB and four other people of high standing in the banking profession. While the Executive Board – as the name suggests – is responsible for *executing* monetary policy, the monetary policy of the ECB is actually designed by the Governing Council, which comprises the whole of the Executive Board plus the governors of the national central banks in the Euro system (22 people in total). The Governing Council, which meets twice a month in Frankfurt, is the most important decision-making body of the ECB and decides, for example, on the level of the ECB's key interest rate, the refinancing rate. The Governing Council also decides how to interpret its duty to achieve price stability. In October 1998 it agreed that price stability should be defined as a 'year-on-year increase in prices of less than 2 per cent as measured by the annual change in a harmonized index of consumer prices throughout the euro area'. A problem with this definition of price stability, however, is that 'less than 2 per cent' is a little vague – an annual inflation rate of 1 per cent and an annual inflation rate of 0 per cent are both less than 2 per cent. In fact, some people were worried that the ECB might even aim for falling prices or negative inflation in order to achieve its target of less than 2 per cent. As we discuss more fully in the coming chapters, this would tend sharply to reduce output and employment in the economy, especially in the short to medium run. In May 2003, therefore, the Governing Council confirmed its official definition thus:

in the pursuit of price stability, it aims to maintain inflation rates below but close to 2 per cent over the medium term.

An important feature of the ECB and of the Euro system in general is its independence. When performing Euro system-related tasks, neither the ECB, nor a national central bank, nor any member of

their decision-making bodies is allowed to seek or take instructions from any external body, including any member governments or any European Union institutions.

The President of the ECB and other members of the Executive Board are appointed for a minimum non-renewable term of office of eight years (although a system of staggered appointments was used for the first Executive Board for members other than the President in order to ensure continuity) and the governors of the 18 national central banks in the Euro system are appointed for a minimum renewable term of office of five years.

THE BANK OF ENGLAND

The **Bank of England** was founded in 1694, although it is not the oldest European central bank (the Swedish Riksbank was founded in 1668). Arguably the most significant event in the Bank of England's 300-year history was when the UK government granted it independence in the setting of interest rates in 1997, which was formalized in an Act of Parliament in 1998. The important body within the Bank that makes the decision on the level at which to set the Bank's key interest rate, the repo rate, is the Monetary Policy Committee (MPC). The MPC consists of the Governor and two Deputy Governors of the Bank of England, two other members appointed by the Bank after consultation with the Chancellor of the Exchequer (the UK finance minister) and four other members appointed by the Chancellor. The Governor and the two Deputy Governors serve five-year renewable terms of office, while other MPC members serve three-year renewable terms. The MPC meets monthly and its interest rate decision is announced immediately after the meeting.

Bank of England the central bank of the United Kingdom

Like the ECB, one of the Bank of England's primary duties is to deliver price stability. Also in common with the ECB, it enjoys independence in the setting of monetary policy – and in particular interest rates – in order to achieve the objective of price stability. Unlike the ECB, however, the Bank of England does not have the freedom to define for itself precisely what 'price stability' means in this context. This is done by the UK government and, in particular, by the Chancellor of the Exchequer. In fact, the 1998 Bank of England Act requires that the Chancellor write to the Governor of the Bank of England once a year to specify what price stability is to be defined as. Currently, the inflation target of 2 per cent is expressed in terms of an annual rate of inflation based on the consumer prices index (CPI). If the target is missed by more than 1 percentage point on either side, i.e. if the annual rate of CPI inflation is more than 3 per cent or less than 1 per cent, the Governor of the Bank of England must write an open letter to the Chancellor explaining the reasons why inflation has increased or fallen to such an extent and what the Bank proposes to do to ensure inflation comes back to the target.

In changes made to financial regulation in the UK in June 2010, the Chancellor of the Exchequer announced that the Bank of England would have new responsibilities focusing on monetary policy and financial stability which have been confirmed in the Financial Services Bill 2012. The new system led to the Bank of England getting additional responsibilities for financial stability, macro-prudential supervision and oversight of micro-prudential supervision. These regulatory functions are overseen by three main groups, the Prudential Regulatory Authority (PRA), which is responsible for day-to-day supervision of bank safety and soundness (micro-prudential policy), the Financial Policy Committee (FPC), charged with 'identifying, monitoring and addressing systemic risks to the UK financial system' and the Financial Conduct Authority (FCA) which manages protection of investors, market supervision and regulation and the conduct of banks and financial services.

SELF TEST If the central bank wants to increase the supply of money, how may it do so?

BANKS AND THE MONEY SUPPLY

So far we have introduced the concept of ‘money’ and discussed how the central bank controls the supply of money by buying and selling government bonds and other assets in open-market operations. Although this explanation of the money supply is correct, it is not complete. In particular, it omits the central role that banks play in the monetary system.

Recall that the amount of money you hold includes both currency (the banknotes in your wallet and coins in your pocket) and demand deposits (the balance in your current account). Because demand deposits are held in banks, the behaviour of banks can influence the quantity of demand deposits in the economy and, therefore, the money supply. This section examines how banks affect the money supply and how they complicate the central bank’s job of controlling the money supply.

The Simple Case of 100 Per Cent Reserve Banking

To see how banks influence the money supply, it’s useful to imagine first a world without any banks at all. In this simple world, currency is the only form of money. To be concrete, let’s suppose that the total quantity of currency is €100. The supply of money is, therefore, €100.

Now suppose that someone opens a bank, appropriately called First European Bank. First European Bank is only a depository institution – that is, it accepts deposits but does not make loans. The purpose of the bank is to give depositors a safe place to keep their money. Whenever a person deposits some money, the bank keeps the money in its vault until the depositor comes to withdraw it or writes a cheque against his or her balance. Deposits that banks have received but have not lent out are called **reserves**. In this imaginary economy, all deposits are held as reserves, so this system is called 100 per cent reserve banking.

reserves deposits that banks have received but have not loaned out

We can express the financial position of First European Bank with a T-account, which is a simplified accounting statement that shows changes in a bank’s assets and liabilities. Here is the T-account for First European Bank if the economy’s entire €100 of money is deposited in the bank:

| FIRST EUROPEAN BANK | | | |
|----------------------------|---------|--------------------|---------|
| Assets | | Liabilities | |
| Reserves | €100.00 | Deposits | €100.00 |

On the left-hand side of the T-account are the bank’s assets of €100 (the reserves it holds in its vaults). On the right-hand side of the T-account are the bank’s liabilities of €100 (the amount it owes to its depositors). Notice that the assets and liabilities of First European Bank exactly balance.

Now consider the money supply in this imaginary economy. Before First European Bank opens, the money supply is the €100 of currency that people are holding. After the bank opens and people deposit their currency, the money supply is the €100 of demand deposits. (There is no longer any currency outstanding, for it is all in the bank vault.) Each deposit in the bank reduces currency and raises demand deposits by exactly the same amount, leaving the money supply unchanged. Thus, *if banks hold all deposits in reserve, banks do not influence the supply of money.*

Money Creation With Fractional-Reserve Banking

Eventually, the bankers at First European Bank may start to reconsider their policy of 100 per cent reserve banking. Leaving all that money sitting idle in their vaults seems unnecessary. Why not use some of it to make loans? Families buying houses, firms building new factories and students at university would

all be happy to pay interest to borrow some of that money for a while. Of course, First European Bank has to keep some reserves so that currency is available if depositors want to make withdrawals. But if the flow of new deposits is roughly the same as the flow of withdrawals, First European needs to keep only a fraction of its deposits in reserve. Thus, First European adopts a system called **fractional-reserve banking**.

fractional-reserve banking a banking system in which banks hold only a fraction of deposits as reserves

The fraction of total deposits that a bank holds as reserves is called the **reserve ratio**. This ratio is determined by a combination of government regulation and bank policy. As we discuss more fully later in the chapter, the European Central Bank place a minimum on the amount of reserves that banks hold, called a *reserve requirement* (although, unusually among central banks, the Bank of England does not impose reserve requirements). In addition, banks may hold reserves above the legal minimum, called *excess reserves*, so they can be more confident that they will not run short of cash. Even when there are no minimum reserve requirements at all (as in the UK), banks will set their own reserve ratio at which they deem it is prudent to operate. For our purpose here, we just take the reserve ratio as given and examine what fractional-reserve banking means for the money supply.

reserve ratio the fraction of deposits that banks hold as reserves

Let's suppose that First European has a reserve ratio of 10 per cent. This means that it keeps 10 per cent of its deposits in reserve and lends out the rest. Now let's look again at the bank's T-account:

| FIRST EUROPEAN BANK | |
|----------------------------|--------------------|
| Assets | Liabilities |
| Reserves €10.00 | Deposits €100.00 |
| Loans €90.00 | |

First European still has €100 in liabilities because making the loans did not alter the bank's obligation to its depositors. But now the bank has two kinds of assets: it has €10 of reserves in its vault, and it has loans of €90. (These loans are liabilities of the people taking out the loans but they are assets of the bank making the loans, because the borrowers will later repay the bank.) In total, First European's assets still equal its liabilities.

Once again consider the supply of money in the economy. Before First European makes any loans, the money supply is the €100 of deposits in the bank. Yet when First European makes these loans, the money supply increases. The depositors still have demand deposits totalling €100, but now the borrowers hold €90 in currency. The money supply (which equals currency plus demand deposits) equals €190. Thus, *when banks hold only a fraction of deposits in reserve, banks create money*.

At first, this creation of money by fractional-reserve banking may seem too good to be true because it appears that the bank has created money out of thin air. To make this creation of money seem less miraculous, note that when First European Bank lends out some of its reserves and creates money, it does not create any wealth (wealth, as a concept, is a 'stock' as opposed to a flow). Loans from First European give the borrowers some currency and thus the ability to buy goods and services. Yet the borrowers are also taking on debts, so the loans do not make them any richer. In other words, as a bank creates the asset of money, it also creates a corresponding liability for its borrowers. At the end of this process of money creation, the economy is more liquid in the sense that there is more of the medium of exchange, but the economy is no wealthier than before.

The Money Multiplier

The creation of money does not stop with First European Bank. Suppose the borrower from First European uses the €90 to buy something from someone who then deposits the currency in Second European Bank. Here is the T-account for Second European Bank:

| SECOND EUROPEAN BANK | | | |
|----------------------|--------|-------------|--------|
| Assets | | Liabilities | |
| Reserves | €9.00 | Deposits | €90.00 |
| Loans | €81.00 | | |

After the deposit, this bank has liabilities of €90. If Second European also has a reserve ratio of 10 per cent, it keeps assets of €9 in reserve and makes €81 in loans. In this way, Second European Bank creates an additional €81 of money. If this €81 is eventually deposited in Third European Bank, which also has a reserve ratio of 10 per cent, this bank keeps €8.10 in reserve and makes €72.90 in loans. Here is the T-account for Third European Bank:

| THIRD EUROPEAN BANK | | | |
|---------------------|--------|-------------|--------|
| Assets | | Liabilities | |
| Reserves | €8.10 | Deposits | €81.00 |
| Loans | €72.90 | | |

The process goes on and on. Each time that money is deposited and a bank loan is made, more money is created. How much money is eventually created in this economy? Let's add it up:

| | |
|-------------------------|----------------------------|
| Original deposit | = €100.00 |
| First European lending | = €90.00 [= 0.9 × €100.00] |
| Second European lending | = €81.00 [= 0.9 × €90.00] |
| Third European lending | = €72.90 [= 0.9 × €81.00] |
| . | . |
| . | . |
| . | . |
| | |
| Total money supply | = €1,000.00 |

It turns out that even though this process of money creation can continue forever, it does not create an infinite amount of money. If you laboriously add the infinite sequence of numbers in the foregoing example, you find the €100 of reserves generates €1,000 of money. The amount of money the banking system generates with each euro of reserves is called the **money multiplier**. In this imaginary economy, where the €100 of reserves generates €1,000 of money, the money multiplier is 10. It is important to remember that there is no new cash being created in this process – banks do not print more notes and press more coins. Most transactions in modern economies are simply 'book entries'; when you get a bank statement telling you that there is a balance of €1,500 in your current account there is not a box with this sum of money stored somewhere in the bank's vault. However, the banking system is such that we have trust that if we did wish to withdraw all that money in cash the bank would have sufficient funds to be able to meet our demand. This is where a 'run on banks' causes so many problems; if a large number of account holders all wanted to withdraw their money as cash the bank is unlikely to have the funds to give to everyone and it would effectively be 'bankrupt'.

money multiplier the amount of money the banking system generates with each unit of reserves

What determines the size of the money multiplier? It turns out that the answer is simple: the money multiplier is the reciprocal of the reserve ratio. If R is the reserve ratio for all banks in the economy, then each euro of reserves generates $1/R$ euros of money. In our example, $R = 1/10$, so the money multiplier is 10.

This reciprocal formula for the money multiplier makes sense. If a bank holds €1,000 in deposits, then a reserve ratio of $1/10$ (10 per cent) means that the bank must hold €100 in reserves. The money multiplier just turns this idea around: if the banking system as a whole holds a total of €100 in reserves, it can have only €1,000 in deposits. In other words, if R is the ratio of reserves to deposits at each bank (that is, the reserve ratio), then the ratio of deposits to reserves in the banking system (that is, the money multiplier) must be $1/R$.

This formula shows how the amount of money banks create depends on the reserve ratio. If the reserve ratio were only $1/20$ (5 per cent), then the banking system would have 20 times as much in deposits as in reserves, implying a money multiplier of 20. Each euro of reserves would generate €20 of money. Similarly, if the reserve ratio were $1/5$ (20 per cent), deposits would be 5 times reserves, the money multiplier would be 5, and each euro of reserves would generate €5 of money. Thus, the higher the reserve ratio, the less of each deposit banks lend out, and the smaller the money multiplier. In the special case of 100 per cent reserve banking, the reserve ratio is 1, the money multiplier is 1, and banks do not make loans or create money.

The Central Bank's Tools of Monetary Control

As we have already discussed, the central bank is responsible for controlling the supply of money in the economy. Now that we understand how fractional-reserve banking works, we are in a better position to understand how the central bank carries out this job. Because banks create money in a system of fractional-reserve banking, the central bank's control of the money supply is indirect. When the central bank decides to change the money supply, it must consider how its actions will work through the banking system.

In general, a central bank has three main tools in its monetary toolbox: open-market operations, the refinancing rate and reserve requirements.

Open-market Operations We introduced the idea of open-market operations earlier through two simple examples. Let's look at these again. If the central bank wants to increase the money supply, it can create currency and use it to buy bonds from the public in the bond market. After the purchase, the extra currency is in the hands of the public. Thus, an open-market purchase of bonds by the central bank increases the money supply. If, on the other hand, the central bank wants to decrease the money supply, it can sell bonds from its portfolio to the public. After the sale, the currency it receives for the bonds is out of the hands of the public. Thus an open-market sale of bonds by the central bank decreases the money supply. To be precise, the open-market operations discussed in these simple examples are called **outright open-market operations**, because they each involve an outright sale or purchase of non-monetary assets to or from the banking sector without a corresponding agreement to reverse the transaction at a later date.

outright open-market operations the outright sale or purchase of non-monetary assets to or from the banking sector by the central bank without a corresponding agreement to reverse the transaction at a later date

The Refinancing Rate The central bank of an economy will set an interest rate at which it is willing to lend to commercial banks on a short-term basis. As we shall see, the name of this interest rate differs across central banks, although in general in this book we'll follow the practice of the European Central Bank and refer to it as the refinancing rate.

The way in which the central bank lends to the banking sector is through a special form of open-market operations. In the previous section we discussed the use of *outright* open-market operations. Although outright open-market operations have traditionally been used by central banks to regulate the money supply, central banks nowadays more often use a slightly more sophisticated form of open-market

operations that involves buying bonds or other assets from banks and at the same time agreeing to sell them back later. When it does this, the central bank has effectively made a loan and taken the bonds or other assets as collateral or security on the loan. The central bank will have a list of eligible assets that it will accept as collateral – ‘safe’ assets such as government bonds or assets issued by large corporations, on which the risk of default by the issuer is negligible. The interest rate that the central bank charges on the loan is the refinancing rate. Because the central bank has bought the assets but the seller has agreed to buy them back later at an agreed price, this kind of open-market operation is often called a **repurchase agreement** or ‘repo’ for short. To see how central bank’s use repos as a means of controlling the money supply and how this is affected by the refinancing rate, we need to look a little more closely at the way commercial banks lend money to one another and borrow from the central bank.

repurchase agreement the sale of a non-monetary asset together with an agreement to repurchase it at a set price at a specified future date

As we discussed earlier, banks need to carry enough reserves to cover their lending and will generally aim for a certain ratio of reserves to deposits, known as the reserve ratio. The minimum reserve ratio may be set by the central bank, but even if it isn’t, banks will still have a reserve ratio that they consider prudent. Now, because deposits and withdrawals at banks can fluctuate randomly, some banks may find that they have an excess of reserves one day (i.e. their reserve ratio is above the level the bank considers prudent or above the minimum reserve ratio, or both), while other banks may find that they are short of reserves and their reserve ratio is too low. Therefore, the commercial banks in an economy will generally lend money to one another on a short-term basis – overnight to a couple of weeks – so that banks with excess reserves can lend them to banks that have inadequate reserves to cover their lending. This market for short-term reserves is called the **money market**. If there is a *general* shortage of liquidity in the money market (because the banks together have done a lot of lending), then the short-term interest rate at which they lend to one another will begin to rise, while it will begin to fall if there is excess liquidity among banks. The central bank closely monitors the money market and may intervene in it in order to affect the supply of liquidity to banks, which in turn affects their lending and hence affects the money supply.

money market the market in which the commercial banks lend money to one another on a short-term basis

Suppose, for example, that there is a shortage of liquidity in the market because the banks have been increasing their lending and they need to increase their reserves. A commercial bank may then attempt to obtain liquidity from the central bank by selling assets to the central bank and at the same time agreeing to purchase them back a short time later. As we said before, in this type of open-market operation the central bank effectively lends money to the bank and takes the assets as collateral on the loan. Because the commercial bank is legally bound to repurchase the assets at a set price, this is called a ‘repurchase agreement’ and the difference between the price the bank sells the assets to the central bank and the price at which it agrees to buy them back, expressed as an annualized percentage of the selling price, is called the repurchase or repo rate by the Bank of England and the refinancing rate by the European Central Bank. The ECB’s **refinancing rate** is thus the rate at which it will lend to the banking sector of the euro area, while the **repo rate** is the rate at which the Bank of England lends short term to the UK banking sector.

refinancing rate the interest rate at which the European Central Bank lends on a short-term basis to the euro area banking sector

repo rate the interest rate at which the Bank of England lends on a short-term basis to the UK banking sector

In the example given, the central bank added liquidity to the banking system by lending reserves to banks. This would have the effect of increasing the money supply. Because the loans made through

open-market operations are typically very short term, with a maturity of at most two weeks, however, the banks are constantly having to repay the loans and borrow again, or 'refinance' the loans. If the central bank wants to mop up liquidity it can simply decide not to renew some of the loans. In practice, however, the central bank will set a reference rate of interest – the Bank of England's repo rate or the ECB's refinancing rate – and will conduct open-market operations, adding to or mopping up liquidity, close to this reference rate.

In the USA the interest rate at which the Federal Reserve lends to the banking sector (corresponding to the ECB's refinancing rate or the Bank of England's repo rate) is called the **discount rate**.

discount rate the interest rate at which the Federal Reserve lends on a short-term basis to the US banking sector

Now we can see why the setting of the central bank's refinancing rate is the key instrument of monetary policy. If the central bank raises the refinancing rate, commercial banks will try and pull in their lending rather than borrow reserves from the central bank, and so the money supply will fall. If the central bank lowers the refinancing rate, banks will feel freer to lend, knowing that they will be able to borrow more cheaply from the central bank in order to meet their reserve requirements, and so the money supply will tend to rise.

Reserve Requirements The central bank may also influence the money supply with **reserve requirements**, which are regulations on the minimum amount of reserves that banks must hold against deposits. Reserve requirements influence how much money the banking system can create with each euro of reserves. An increase in reserve requirements means that banks must hold more reserves and, therefore, can lend out less of each euro that is deposited; as a result, it raises the reserve ratio, lowers the money multiplier and decreases the money supply. Conversely, a decrease in reserve requirements lowers the reserve ratio, raises the money multiplier and increases the money supply.

reserve requirements regulations on the minimum amount of reserves that banks must hold against deposits

Central banks have traditionally tended to use changes in reserve requirements only rarely because frequent changes would disrupt the business of banking. When the central bank increases reserve requirements, for instance, some banks find themselves short of reserves, even though they have seen no change in deposits. As a result, they have to curtail lending until they build their level of reserves to the new required level.

As we have noted, the Bank of England no longer sets minimum reserve requirements but the European Central Bank does set minimum reserve requirements. The ECB applies reserves to the average reserve ratio over a specified period rather than at a single point in time. It does this to stop the amount of lending fluctuating wildly, in order to maintain stability in the money market. Hence, the ECB uses reserve requirements in order to maintain stability in the money market rather than as an instrument of policy by which to increase or decrease the money supply. Following the financial crisis, negotiations have taken place on improving banks' reserves to avoid the problems faced during the crisis. The so-called Basel III negotiations between 27 countries set new reserve requirements in September 2010. The new rules came into force in 2013 and are to be phased in over a period of six years (although some relaxation to the requirements on the capital reserves needed to survive a 30-day crisis was agreed in 2013 and will be phased in from 2015). The regulations mean that banks will have to have higher reserves to support lending; for every €50 of lending banks will have to have €3.50 of reserves compared to €1 prior to the Basel III agreement. This obviously more than triples the amount of reserves that banks will have to keep. If banks do not adhere to the new regulations then they risk seeing the authorities placing restrictions on their activities including paying out dividends to shareholders and bonuses to staff.

Quantitative Easing

During the financial crisis of 2007–2009, central banks adopted new tactics to try and support the economy, which involve market operations. One of the methods adopted was an asset purchasing facility (APF) or quantitative easing (QE). In the UK the Bank of England cut interest rates to 0.5 per cent and in the US rates fell to a target of between 0 and 0.25 per cent. Having effectively exhausted the use of lowering the price of money in the economy, central banks have looked at affecting the supply of money as part of their armoury in triggering economic activity. QE was the method adopted by the UK (authority to carry out QE was given in January 2009, although the first purchase was 11 March 2009) and the US (from March 2009).

All banks have accounts at the central bank, which are classed as ‘reserves’. The process of QE involves the central bank buying assets from private sector institutions. These might include banks, pension funds and insurance companies. The type of assets purchased varies from bonds and gilt-edged stock to commercial paper (short-term promissory notes issued by companies with a maturity ranging up to 270 days but with an average of 30 days), equities and possibly toxic assets. When the central bank buys assets, these reserves increase.

In buying these assets, the central bank is effectively increasing the amount of money in the accounts of those who sell these assets. The purpose of the policy is to encourage banks and other financial institutions, who are the recipients of this additional money, to lend it out to businesses and consumers thus stimulating demand. An example may serve to help understand the process (we will use pounds in this example and assume the central bank is the Bank of England). Assume a bank has £1 billion worth of gilt-edged stock. It decides that it wants to sell £500 million worth. The Bank of England announces that it intends to purchase gilts at an auction and sets a specific date for the auction (normally a Monday and Wednesday each week for a specified period of time). The process takes place via what is called a ‘reverse auction’. Rather than the buyer putting in bids for what they would be willing to pay for the item, the seller submits an electronic bid to the Bank stating what they would like to sell and at what price they would be willing to sell at. The Bank itself has a specified amount that it intends to buy at that time and the bids coming in from banks and other institutions may mean that the auction is oversubscribed (i.e. more bids to sell are received than the Bank intends to buy) and as a result the Bank is able to select which offers it will accept and at what price.

Assume our bank has been successful in its bid. It transfers the gilts to the Bank of England and in return receives a credit for £500 million. It can use this £500 million to lend to businesses for investment into equipment and machinery and consumers for mortgages or personal loans for cars, students loans, new furniture and so on. The additional spending on investment and consumer goods helps boost aggregate demand and so trigger economic growth. Given that the central bank is actively participating in the bond market we can also expect some effects through this market action. Remember, companies use the bond market as a means of raising funds. Over a period of time these bonds will mature and firms may issue new bonds to replace them. These new bond issues may have different coupon rates to existing ones depending on current conditions in the bond market. Assume that a company has issued a ten-year bond with a coupon, on issue, of 5 per cent. The price at which the bond sells on the market will not necessarily be the same as its par value (what it was originally issued at). The ratio of coupon in relation to the price gives investors the yield $\left(Yield = \frac{Coupon}{Price} \right)$. For example, if the company issues a €100 bond, with a coupon (interest) rate of 5 per cent, the yield will also be 5 per cent. However, if demand for this bond rises then its price may rise (to €105 for example) and as a result the yield will fall:

$$\left(\frac{5 \text{ per cent}}{105} = 4.76 \text{ per cent} \right).$$

This is the inverse relationship between bond prices and yield we introduced in an earlier chapter.

If the central bank intervenes to buy bonds the supply of bonds will fall and bond prices will rise. As bond prices rise, yields fall. If our company now issues a new bond it can offer that bond at a lower coupon (there is no incentive not to buy it since buying existing bonds does not give any better return). Using the example above, if our company wanted to issue a new bond to replace the matured €100 bond, then it

could offer the bond at a coupon of 4.76 per cent (or slightly higher) and have every chance of raising the finance. For the company it is now raising funds at 4.76 per cent rather than 5 per cent. This means that firms are now able to borrow at cheaper rates. The rise in price of gilts and subsequent fall in yields reduces the spread (the difference between the interest rate on borrowed money and the return on the gilt). As the spread falls there is more incentive for the bank to look for improved returns elsewhere and it is hoped that it will consider lending to consumers and businesses (who may have been seen as being too risky previously) because the return is now greater.

QE, therefore, works in a variety of ways to influence incentives as highlighted in Principle 4 of the *Ten Principles of Economics*. In essence, therefore, the principle of QE is simple; the acid test is whether it works or not. The central bank is in a position to monitor the effects of the process as it collects data about money flows in the economy, the effect on the money supply, credit flows in corporate markets, interest rates on different types of lending (for example on mortgages) and the amount and type of lending taking place.

The type of debt bought by the central bank is critical to the success of the policy. Given the situation that many banks found themselves in as a result of the financial crisis, it was possible that if the central bank focused its purchases of assets from these institutions, then rather than lending out the money again and freeing up credit markets, banks may simply have used the receipts to build up their balance sheets. As a result the effect would be weaker than planned.

The Bank of England has admitted that it is difficult to know whether QE is working or how long it would take to see any measurable effects. One of the problems is something called the *counterfactual* – what the situation would have been if QE was not implemented. By the end of December 2012, the Bank of England had spent £375 billion on QE. If it had not spent this £375 billion on QE what state would the economy be in? Would it have been in a much worse state and if so how much worse?

Problems in Controlling the Money Supply

Through the setting of its refinancing rate and the associated open-market operations, the central bank can exert an important degree of control over the money supply. Yet the central bank's control of the money supply is not precise. The central bank must wrestle with two problems, each of which arises because much of the money supply is created by the system of fractional-reserve banking.

The first problem is that the central bank does not control the amount of money that households choose to hold as deposits in banks. The more money households deposit, the more reserves banks have, and the more money the banking system can create. And the less money households deposit, the less reserves banks have, and the less money the banking system can create. To see why this is a problem, suppose that one day people begin to lose confidence in the banking system and, therefore, decide to withdraw deposits and hold more currency. When this happens, the banking system loses reserves and creates less money. The money supply falls, even without any central bank action.

The second problem of monetary control is that the central bank does not control the amount that bankers choose to lend. When money is deposited in a bank, it creates more money only when the bank lends it out. Because banks can choose to hold excess reserves instead, the central bank cannot be sure how much money the banking system will create. For instance, suppose that one day bankers become more cautious about economic conditions and decide to make fewer loans and hold greater reserves. In this case, the banking system creates less money than it otherwise would. Because of the bankers' decision, the money supply falls.

Hence, in a system of fractional-reserve banking, the amount of money in the economy depends in part on the behaviour of depositors and bankers. Because the central bank cannot control or perfectly predict this behaviour, it cannot perfectly control the money supply.

SELF TEST Describe how banks create money. • If the ECB wanted to use three of its policy tools to decrease the money supply, what would it do?

Changes in Banking and the Financial Crisis

A traditional view of banking is closely allied with the idea of prudence. Being prudent may make for safe banking but it does not necessarily satisfy the demands of shareholders for ever larger profits. In the latter part of the 1990s and through to 2008, banks broadened the scope of their activities to generate higher profits. The financial crisis of 2007–2009 was due, in part, to banks taking on riskier activities and an increasingly blurred demarcation from a bank's retail (their dealings with members of the public) and wholesale operations (where the bank deals with corporate and other financial institutions). We have seen how financial institutions used 'quants' to try and reduce risk. This was part of the process that led to the growth of sub-prime lending. With the benefit of hindsight it might seem rather foolish for anyone, let alone a bank, to have contemplated lending money to people with bad credit histories and low incomes.

The growth in the housing market during this period along with the changes brought about by deregulation of financial markets in North America, Europe and the UK created the circumstances which allowed banks and other mortgage lenders to expand the amount of loans they made. Under 'traditional' lending regimes banks would have had to have set aside reserves to cover the loans that they had made. That assumes, of course, that such loans appeared on the bank's balance sheet. Innovation in the financial services industry and deregulation coupled with some ingenious thinking by the quants led to ways being developed which allowed banks to lend funds without it appearing on their balance sheets and as a result they were able to offload the risk to others as well as allow more and more lending to be possible.

In our example of reserve banking, reserves were assumed to be 10 per cent. Banks do not simply wait for deposits to be made and then use those deposits as reserves against loan multiples. Banks want to make loans as this is one key way in which they can make profits. If they want to increase lending, therefore, they have to attract sufficient deposits to provide the reserves needed. For example, assume a bank wants to increase lending by €1 billion. It knows that it has to have cash reserves to cover such lending of €100 million (10 per cent of €1 billion). If it set reserves lower, say at 8 per cent, then it would only have to attract €80 million in deposits. The bank therefore has to manage the risks inherent in its balance sheet. It has to make sure that any loan defaults are not more than the amount it has in reserve. Under normal circumstances it will have to market its services and products aggressively in order to attract deposits of €100 million. In doing so it will have to offer more attractive interest rates and this will cut the profitability of its loans.

However, if the bank can make €1 billion of loans without the loans appearing on its balance sheet then it does not have to attract the deposits to act as reserves. These off-balance sheet loans attract income in the form of interest and fees. The result is a greater capacity for lending whilst increasing potential profits for the bank. Banks found that they could put together collections of debt and sell them on. These parcels of debt did have risk attached to them but this risk could be reduced by taking out credit default swaps. This securitization of assets represented one of the innovations in banking that grew with the expansion of sub-prime lending.

Securitization of Assets The basis of the process lay in the fact that the assets involved generated cash flows over a period of time. Securitization takes loans off the balance sheet and so the bank does not have to set aside reserves to cover those loans. This means that the bank has more scope for increasing lending. Assume a bank has agreed to lend 1,000 individuals mortgages for property with each individual borrowing €100,000. The bank is known as the *originator*. These mortgages could be bundled up to form €1 billion worth of debt. Mathematicians analyse the debt and make an assessment that the debt has a 1 per cent chance of default. That is to say, of the 1,000 individuals associated with this debt, ten are likely to default at some point over the lifetime of the debt. The bank can present the debt package to a credit rating agency and, given the limited risk involved as a result of the pooling of the debt, is able to access a favourable credit rating.

Having secured an appropriate credit rating for the package of debt the bank then sets up what is called a *special purpose vehicle* or SPV. The establishment of an SPV allows the bank to separate its financial obligations. Rather than buyers of the debt having a claim on the bank as a whole in case of default, the setting up of an SPV means that the investor has a claim against the SPV but not the bank. Equally, the investor has the right to receive payments first by dealing with an SPV. The SPV buys the collection of

debt and will sell it to investors which may be other banks and financial institutions. The funds raised by the SPV allow it to be able to buy the debt from the bank. The SPV will issue bonds for this purpose. For the prospective investor the bond is associated with the stream of cash flows from the package of loans which also have a relatively high rate of interest attached to them. Given that the debt package has been given a high credit rating, the risk associated with the investment is considered relatively low. The investor is protected by the value of the underlying assets. If there was a default on any of the loans the bank had the option of seizing the properties which are the security for the original loan.

The SPV will issue bonds for a total value lower than the package of debt. The difference is the first-loss position of the originator (the bank in this example). The aim is to ensure that the potential loss on the pool of debt is not greater than the difference between the total value of the pool and the value of the bonds issued to sell on the debt. In our example, the pool of debt, the collection of mortgages, is worth €1 billion. The SPV will receive the mortgage payments and uses the cash flows to pay the bond interest and the principal. The debt is sold by the SPV for €950 million – the difference, therefore being €50 million. Analysis shows that the likely default losses on this package of debt would be no more than €30 million. This means that the cash inflows from mortgage owners is sufficient to pay the investors their money back – €950 million. This represents a 3 per cent loss rate based on the total value of the debt package.

The SPV issues shares in itself and, therefore, becomes a subsidiary company of the bank (or originator) with its own legal status. Shares may be bought by the originator and by other parties in the financial community. The attraction of doing so is that the SPV generates high returns based on the flow of cash of the underlying asset (the mortgage). The benefit of setting up an SPV for the bank is that it transfers liabilities to the SPV and therefore these do not appear on its balance sheet. As a result the bank does not have to set aside reserves to cover these loans and this leaves it free to increase lending or engage in deals which further expand its earnings and profits. Such activity may include providing finance for acquisitions or making loans to hedge funds. The setting up of an SPV also protects investors. If the bank (which originated the debt) fails, then creditors of the bank (those to whom the bank owes money) cannot make any claim against investors given that the SPV is a separate legal entity. Equally, if the SPV fails in some way then the investor does not have recourse to any claim against the bank or the originator. The bank can also earn commission on the sale of the debt to the SPV.

The issue of bonds by the SPV would be in bundles, called *tranches*, which reflect different levels of risk. Given these different levels of risk, different types of investors would be attracted for any particular tranche. Each tranche might relate to the maturity date of the mortgages in the pool or be associated with a different rate of interest, but typically in sub-prime securitization there were six tranches (known as '6-pack deals'). Each tranche can be sold separately and much of this bundled debt could be 'insured' using CDS.

The global nature of the financial system meant that what was created was a complex web of financial transactions that built high levels of interdependence among those involved but where few had any real knowledge of the extent of this interdependence. In essence, the whole structure of mortgage lending was dependent on the ability of those who had taken out mortgages (including sub-prime) to be able to pay the principal and interest on their loans. When the number of borrowers defaulting on their repayments started to rise in the middle of the 2000s, the structure began to become unstable and the full extent of the interdependence began to reveal itself.

Mortgages with rates that can change are referred to as variable rate or adjustable rate mortgages. Most of these mortgages are linked not simply to central bank base rate but to 'LIBOR', the London Interbank Offered Rate, which is a benchmark set of interest rates which determine other interest rates in the economy and is the rate at which banks will lend to each other in the London money markets. In the United States around \$350 trillion of loans are based on LIBOR.

Each morning the British Bankers' Association gets quotes from 16 major banks on the rates at which they are prepared to borrow ten currencies with 15 different maturities. These maturities are the spot/overnight next (i.e. for loans made today and repaid tomorrow), one week, two weeks, one month, two months and so on up to a maximum of 12 months. Whilst ten currencies are quoted, the four main ones are the dollar, British pound, euro and Japanese yen.

The quotes are ranked in order of size and an average of the middle two quantities taken to give the LIBOR for that currency. The details are published at 11.30 a.m. every morning and form the basis of short-term lending. If LIBOR rates rise then adjustable rate mortgages that are tied to it also rise. Some

73 per cent of adjustable rate sub-prime mortgages in the US were based on six-month LIBOR and not the Federal Funds rate in the period April–September 2007.

From around 2005, central banks began pushing up interest rates to counter inflationary pressures. As interest rates rose, borrowers, especially those on sub-prime mortgages, began to feel the pressure and reports of the number of defaults on sub-prime loans began to rise. Once mortgage defaults began to rise faster the number of financial institutions affected and their exposure became clearer. The whole edifice was based on the expectation that the underlying assets would continue to generate the income stream over time – in other words, that the very large majority of mortgage holders would continue to pay their monthly mortgage repayments. As mortgage rates rose borrowers found it increasingly difficult to meet their monthly payments.

If mortgage payers default on loans then banks who lent money to SPVs can call in that debt. Banks who set up the SPV may then have to take the assets of the SPV back onto its balance sheet. This not only limits their ability to lend further because they now have to put aside reserves to cover these liabilities which are now on their balance sheet, but they may also have to write down the value of the assets, further damaging their balance sheet and ability to lend. Every day, banks have obligations to meet – loans they have taken out that need paying off, interest payments, CDS that mature, bonds that need to be paid, etc. They must have sufficient liquid funds to be able to meet these obligations. Many banks borrow the funds they need from the interbank market. As the sub-prime market collapsed, the exposure to bad debt started to become more obvious and a number of banks reported significant write-downs and losses. Confidence in the banking system, so important to its functioning, began to fall. Banks were not sure of their own exposure to these bad debts (referred to as **toxic debt**) and so were also unsure about the extent of other banks' exposure. Interbank lending began to become much tighter as banks were unwilling to lend to each other and they also faced the task of trying to shore up their own balance sheets.

toxic debt mortgage-backed securities and other debt (such as bonds) that are not able to be repaid in many cases because the value of the assets against which they are secured have fallen significantly

A significant proportion of the funds that banks borrow are from each other. As mentioned above, the rate at which they borrow from each other is LIBOR. Under normal circumstances, the difference between bank rate and LIBOR is small and stable. In times when credit is scarce, the supply of loans falls and as a result the price rises. Between March 2008 and October 2008 LIBOR rose sharply. Central banks around the world responded to the tightness in the credit market, now termed the 'credit crunch', by cutting interest rates and injecting funds in to the markets. It took until June 2008 for LIBOR gradually to fall back to the Fed Funds level, reflecting the continued tightness in the credit market. Accessing credit was, therefore, far more expensive and limited in nature and the higher price of borrowing was passed onto the consumer – individuals and businesses. As business loans are less easy to obtain or more expensive, businesses find it difficult to manage their cash flow and insolvency can result.

Around this time, regulators began investigating reports of irregularities in the way in which LIBOR was being set. Subsequent investigations uncovered that traders at a number of banks had systematically attempted to fix LIBOR rates. Rates were being set artificially low with the result being that the true cost of money in the markets was not being reflected in the interest rate. The scandal raised serious issues for the necessity of trust in the banking and financial system so that decisions can be based on accurate market data. In the wake of the investigations a number of banks were fined, senior executives of some banks resigned and legal action is being taken against traders involved in the scandal.

The Crisis Unfolds In August 2008, the problems facing banks began to mount – how far the attempts to fix LIBOR contributed to the problems is still being debated. However, in March 2008, the US investment bank Bear Stearns' exposure to sub-prime markets led it to seek support from the US Federal Reserve (the Fed); a year before, New Century Financial had sought protective bankruptcy and needed financial assistance. Other banks with a heavy exposure to sub-prime began to announce losses and write-downs only to have to issue worse revised figures a short time later. Bear Stearns was rescued by being acquired by J.P. Morgan Chase. In the UK, Northern Rock suffered a run on its assets as account holders queued to

get their money before the bank collapsed – the first run on a UK bank for more than a hundred years. The UK government had to step in to take Northern Rock into public hands and nationalize it.

In the US, it became obvious that two of the major players in the mortgage market: the Federal National Mortgage Association, commonly known as 'Fannie Mae'; and the Federal Home Loan Mortgage Corporation, known as 'Freddie Mac', were facing deep financial problems. Their business involved buying mortgages from lenders and then selling on the debt to investors. They effectively guaranteed the borrowing for millions of mortgage owners and accounted for around half of the US mortgage market which was worth \$12 trillion. Such was their importance that the US authorities stepped in to support them. As in the UK with Northern Rock, the US government announced in early September 2008 that it was going to take temporary ownership of the two companies to save them from collapse. If the two had become insolvent, house prices in the US, which were falling at rates of around 15 per cent in some areas, would have been likely to have fallen even further, and a larger number of people would have fallen into negative equity. It was estimated at the time that up to around 10 per cent of US homeowners were facing difficulties in meeting their mortgage payments, and risked having their homes repossessed.

CONCLUSION

In *Le Bourgeois Gentilhomme*, a play by the French playwright Molière, a character called Monsieur Jourdain finds out what prose is and then exclaims, 'Good heavens! For more than 40 years I have been speaking prose without knowing it!' In the same way, newcomers to economics may sometimes be surprised to find that they have been participating in the monetary system of their economy for years without realizing it. Ever since you got your first pocket money you have been using fiat money. Ever since you got your first bank account you have been participating in fractional-reserve banking and the creation of money. Whenever we buy or sell anything, we are relying on the extraordinarily useful social convention called 'money'. Now that we know what money is and what determines its supply, we can discuss how changes in the quantity of money affect the economy. We begin to address that topic in the next chapter.

IN THE NEWS



Open Market Operations

Whilst it is difficult for central banks to control the money supply in precise ways, the use of open market operations can be a useful tool in managing relatively short-term interest rates.

China's Central Bank in Record Market Intervention

In February 2013, it was announced that The People's Bank of China (PBOC), the country's central bank, had engaged in open market operations creating a record in the process. The PBOC bought 450 billion yuan (€54 billion) worth of assets which represented the largest single day injection into the money markets beating the previous record of 395 billion yuan (€47.4 billion) in October 2012. The operation involved the purchase of 14-day reverse bond repurchase agreements. The PBOC purchased securities from the money markets with the agreement to sell them back at a higher price in 14 days (this is the opposite to a repo and so for money market institutions selling the securities

and buying them back in 14 days it is a repo). In the first instance, therefore, the PBOC had injected funds into the money markets but sucked out more funds when it sold the securities back.

Part of the reason for doing this was to meet liquidity needs in the markets prior to the Spring Festival Holiday period in China. Around this time, Chinese citizens tend to withdraw more cash to spend on gifts, food, travel and entertainment for the holiday period. The open market operations will provide the liquidity to be able to cope with this demand rather than having to adjust bank reserve requirements. Chinese officials expressed some concern that cutting reserve requirements for banks might stimulate inflation and so using open market operations in this way helps to reduce the threat of rising inflation and

keeps interest rates under control. If repo rates are rising, it is a sign that demand for money is higher than supply and in the past, observers have noted that there were spikes in repo rates when holiday periods approached or as the month or quarter approached. By using open market operations in this way, the PBOC sent a signal to the markets that it intends to manage interest rates. When the demand for money increases as during holiday periods, the injection of funds into the system helps reduce the pressure on interest rates to rise and when demand falls back, the selling back of the securities reduces the supply of funds and thus maintains interest rates at a more stable level than without the operation.

The People's Bank of China, the central bank of China is likely to become an increasingly important player in global financial markets as China's economy continues to grow.



Questions

- 1 What is meant by the term open market operations?
- 2 What is a repurchase agreement and why was the action of the PBOC termed a 'reverse repurchase agreement'?
- 3 Assume no intervention in the money markets by the PBOC. Use a supply and demand diagram with interest rates on the vertical axis (the price of money) to show how spikes in interest rates can occur in times when the demand for money rises.
- 4 Why might Chinese officials be concerned that a cut in reserve requirements cause inflation?
- 5 Again, using a supply and demand diagram, explain how the open market operations of the PBOC can help to stabilize interest rates.

SUMMARY

- The term *money* refers to assets that people regularly use to buy goods and services.
- Money serves three functions. As a medium of exchange, it provides the item used to make transactions. As a unit of account, it provides the way in which prices and other economic values are recorded. As a store of value, it provides a way of transferring purchasing power from the present to the future.
- Commodity money, such as gold, is money that has intrinsic value: it would be valued even if it were not used as money. Fiat money, such as paper euros or pounds, is money without intrinsic value: it would be worthless if it were not used as money.
- In an advanced economy, money takes the form of currency and various types of bank deposits, such as current accounts.
- A central bank is an institution designed to regulate the quantity of money in an economy.
- The European Central Bank is the overall central bank for the 18 countries participating in European Monetary Union. The Euro system is made up of the European Central Bank plus the corresponding 18 national central banks.
- The UK central bank is the Bank of England. It was granted independence in the setting of interest rates in 1997.
- Central banks control the money supply primarily through the refinancing rate and the associated open-market operations. An increase in the refinancing rate means that it is more expensive for banks to borrow from the central bank on a short-term basis if they are short of reserves to cover their lending, and so they will tend to reduce their lending and the money supply will contract. Conversely, a reduction in the refinancing rate will tend to expand the money supply.
- The central bank can also use outright open-market operations to affect the money supply: a purchase of government bonds and other assets from the banking sector increases the money supply, and the sale of assets decreases the money supply. The central bank can also expand the money supply by lowering minimum reserve requirements and it can contract the money supply by raising minimum reserve requirements.
- When banks lend out some of their deposits, they increase the quantity of money in the economy. Because banks may choose to hold excess reserves, they can affect the supply of money beyond the control of the central bank. When households deposit money in banks, banks can use these deposits to create money. But the central bank cannot control exactly how much money that households wish to deposit. Because of these two factors, the central bank's control of the money supply is imperfect.
- Central banks adopted quantitative easing as a means of easing the financial crisis and stimulating the economy through buying assets from the banking sector and expanding the money supply.
- A bank run occurs when depositors suspect that a bank may go bankrupt and, therefore, 'run' to the bank to withdraw their deposits. Many countries have a system of deposit insurance and central banks are lenders of last resort so bank runs can be managed more effectively.
- The financial crisis revealed the extent to which banking has changed in the last 30 years and in response to the crisis, governments have attempted to impose more stringent regulations on banks.

QUESTIONS FOR REVIEW

- 1 What distinguishes money from other assets in the economy?
- 2 What is commodity money? What is fiat money? Which kind do we use?
- 3 What are demand deposits, and why should they be included in the stock of money?
- 4 Who is responsible for setting monetary policy at the European Central Bank? How is this group chosen? Who is responsible for setting monetary policy at the Bank of England? How is this group chosen?
- 5 If the central bank wants to increase the money supply with outright open-market operations, what does it do?
- 6 Why don't banks hold 100 per cent reserves? How is the amount of reserves banks hold related to the amount of money the banking system creates?
- 7 What is the refinancing rate? What happens to the money supply when the European Central Bank raises its refinancing rate?
- 8 What are reserve requirements? What happens to the money supply when the ECB raises reserve requirements?
- 9 What is quantitative easing and in what way does it help to counter slow economic growth in an economy?
- 10 Why can't central banks control the money supply perfectly?

PROBLEMS AND APPLICATIONS

- 1 What characteristics of an asset make it useful as a medium of exchange? As a store of value?
- 2 Suppose that someone in one of the euro area countries discovered an easy way to counterfeit €100 banknotes. How would this development affect the monetary system of the euro area? Explain.
- 3 Your uncle repays a €100 loan from Tenth European Bank (TEB) by writing a €100 cheque from his TEB current account. Use T-accounts to show the effect of this transaction on your uncle and on TEB. Has your uncle's wealth changed? Explain.
- 4 Beleaguered European Bank (BEB) holds €250 million in deposits and maintains a reserve ratio of 10 per cent.
 - a. Show a T-account for BEB.
 - b. Now suppose that BEB's largest depositor withdraws €10 million in cash from her account. If BEB decides to restore its reserve ratio by reducing the amount of loans outstanding, show its new T-account.
 - c. Explain what effect BEB's action will have on other banks.
 - d. Why might it be difficult for BEB to take the action described in part (b)? Discuss another way for BEB to return to its original reserve ratio.
- 5 You take €100 you had kept under your pillow and deposit it in your bank account. If this €100 stays in the banking system as reserves and if banks hold reserves equal to 10 per cent of deposits, by how much does the total amount of deposits in the banking system increase? By how much does the money supply increase?
- 6 The European Central Bank conducts a €10 million open market purchase of eligible assets from the banking sector. If the required reserve ratio is 10 per cent, what is the largest possible increase in the money supply that could result? Explain. What is the smallest possible increase? Explain.
- 7 Suppose that the T-account for First European Bank is as follows:

| Assets | | Liabilities | |
|----------|----------|-------------|----------|
| Reserves | €100,000 | Deposits | €500,000 |
| Loans | €400,000 | | |

- a. If the ECB requires banks to hold 5 per cent of deposits as reserves, how much in excess reserves does First European now hold?

- b. Assume that all other banks hold only the required amount of reserves. If First European decides to reduce its reserves to only the required amount, by how much would the economy's money supply increase?
- 8 Suppose that the reserve requirement for current deposits is 10 per cent and that banks do not hold any excess reserves.
 - a. If the ECB decides not to renew €1 million of loans it previously made to the euro area banking sector, what is the effect on the economy's reserves and money supply?
 - b. Now suppose the ECB lowers the reserve requirement to 5 per cent, but banks choose to hold another 5 per cent of deposits as excess reserves. Why might banks do so? What is the overall change in the money multiplier and the money supply as a result of these actions?
 - 9 The economy of Elmendyn contains 2,000 €1 coins.
 - a. If people hold all money as currency, what is the quantity of money?
 - b. If people hold all money as demand deposits, and banks maintain 100 per cent reserves, what is the quantity of money?
 - c. If people hold equal amounts of currency and demand deposits, and banks maintain 100 per cent reserves, what is the quantity of money?
 - d. If people hold all money as demand deposits, and banks maintain a reserve ratio of 10 per cent, what is the quantity of money?
 - e. If people hold equal amounts of currency and demand deposits, and banks maintain a reserve ratio of 10 per cent, what is the quantity of money?
 - 10 A bank bundles up €1 billion of mortgage debt with a forecast default ratio of 5 per cent. The debt is sold via an SPV to a group of investors for €800 million.
 - a. The group of investors wish to reduce their risk. Explain how a credit default swap with an insurance company might help them to do this.
 - b. There is a credit event and the original debt suffers a default ratio of 50 per cent. Explain how the originator and the investors might be affected by such an event.

27

MONEY GROWTH AND INFLATION

In 1930, the Lyons Maid company produced choc ices – which they described as chocolate-coated vanilla and coffee flavoured ice-cream bars – which were sold in UK confectionery shops for two old pence. Given that before the UK currency was decimalized in 1970 there were 240 old pence to the pound, this means that £1 would have bought you 120 choc ices in 1927. Today, a Wall's 'Chunky' choc ice – a chocolate-coated vanilla flavoured ice-cream bar – is priced at about 70p, so £1 today will buy you only 1.4 choc ices. This represents quite a substantial price rise.

You are probably not surprised at the increase in the price of ice cream. In advanced economies, most prices tend to rise over time. This increase in the overall level of prices is called *inflation*. Inflation may seem natural and inevitable to a person who has grown up in an advanced economy in western Europe or North America at the end of the 20th century, but in fact it is not inevitable at all. There were long periods in the 19th century during which, in some economies, most prices fell – a phenomenon called *deflation*.

Although inflation has been the norm in more recent history, there has been substantial variation in the rate at which prices rise. Inflation in the UK during the late 1990s and the first half of the 2000s was low and stable at round 2 per cent or so. However, in the mid-1970s, annual UK inflation, as measured by increases in the retail prices index, exceeded 20 per cent.

International data show an even broader range of inflation experiences. Germany after World War I experienced a spectacular example of inflation. The price of a newspaper rose from 0.3 marks in January 1921 to 70,000,000 marks less than two years later. Other prices rose by similar amounts. An extraordinarily high rate of inflation such as this is called *hyperinflation*. The German hyperinflation had such an adverse effect on the German economy that it is often viewed as one contributor to the rise to power of the National Socialists (Nazis) and, as a result, World War II. Over the past 50 years, with this episode still in mind, German policymakers have been extraordinarily averse to inflation, and Germany has had much lower inflation than most other countries of the world.

In more recent times there have been episodes of hyperinflation in the former Yugoslavia and in Zimbabwe. Inflation in Yugoslavia ran at 5 quadrillion per cent (5 with 15 zeros after it) between October 1993 and January 1995. The Zimbabwean authorities announced that inflation had reached 231,000,000 per cent in June 2008. However, the prices of some goods rose by considerably more. The Zimbabwean central bank reported that some goods on the black market had risen by 70,000,000 per cent. Laundry soap was one of the goods that had risen by this much, but cooking oil also rose by 60,000,000 per cent and sugar by 36,000,000 per cent. Inflation made almost every Zimbabwean a billionaire. Unskilled workers earned around Z\$200,000,000,000 a month – at the time, equivalent to about US\$10. In July 2008 the government issued a Z\$100 billion note. If you had one it would just about have bought you a loaf of bread.

What determines whether an economy experiences inflation and, if so, how much? This chapter answers the question by developing the quantity theory of money. Chapter 1 summarized this theory as one of the *Ten Principles of Economics*: prices rise when the government prints too much money. This insight has a long and venerable tradition among economists. The quantity theory was discussed by the famous 18th-century British philosopher and economist, David Hume, and was advocated in the latter part of the 20th century by the prominent American economist Milton Friedman. This theory of inflation can explain both moderate inflations, such as those experienced in the United States, and hyperinflations, such as those outlined above.

After developing a theory of inflation, we turn to a related question: why is inflation a problem? At first glance, the answer to this question may seem obvious: inflation is a problem because people don't like it. In the 1970s, when the United Kingdom (along with many other economies) was experiencing a relatively high rate of inflation, opinion polls placed inflation as the most important issue facing the nation.

But what, exactly, are the costs that inflation imposes on a society? The answer may surprise you. Identifying the various costs of inflation is not as straightforward as it first appears. As a result, although all economists decry hyperinflation, some economists argue that the costs of moderate inflation are not nearly as large as the general public believes.

THE CLASSICAL THEORY OF INFLATION

We begin our study of inflation by developing the quantity theory of money. This theory is often called 'classical' because it was developed by some of the earliest thinkers about economic issues back in the 18th century such as David Hume, who are often referred to as the 'classical economists'. Most economists today rely on this theory to explain the long-run determinants of the price level and the inflation rate.

The Level of Prices and the Value of Money

Suppose over some period of time we observe a ten-fold increase in the price of an ice cream. What conclusion should we draw from the fact that people are willing to give up so much more money in exchange for an ice cream? It is possible that people have come to enjoy ice cream more. Yet, even if people's enjoyment of ice cream has increased, a large amount of the price rise is probably due to the fact that, over time, the money used to buy ice cream has become less valuable. Indeed, the first insight about inflation is that it is more about the value of money (the goods and services any given amount of money can be exchanged for) than about the value of goods.

This insight helps point the way towards a theory of inflation. When the consumer prices index (CPI) and other measures of the price level rise, commentators are often tempted to look at the many individual prices that make up these price indices: 'The CPI rose by 3 per cent last month, led by a 20 per cent rise in the price of coffee and a 27 per cent rise in the price of electricity.' Although this approach does contain some interesting information about what's happening in the economy, it also misses a key point: inflation is an economy-wide phenomenon that concerns, first and foremost, the value of the economy's medium of exchange.

The economy's overall price level can be viewed in two ways. So far, we have viewed the price level as the price of a basket of goods and services. When the price level rises, people have to pay more for the goods and services they buy. Alternatively, we can view the price level as a measure of the value of money. A rise in the price level means a lower value of money because each unit of money now buys a smaller quantity of goods and services.

It may help to express these ideas mathematically. Suppose P is the price level as measured, for instance, by the CPI or the GDP deflator. Then P measures the number of euros needed to buy a basket of goods and services. Now turn this idea around: the quantity of goods and services that can be bought with €1 equals $1/P$. In other words, if P is the price of goods and services measured in terms of money, $1/P$ is the value of money measured in terms of goods and services. Thus, when the overall price level rises, the value of money falls. We gave an example of this in relation to choc ices above.

Money Supply, Money Demand and Monetary Equilibrium

What determines the value of money? The answer to this question, like many in economics, is supply and demand. Just as the supply and demand for bananas determines the price of bananas, the supply and demand for money determines the value of money. Thus, our next step in developing the quantity theory of money is to consider the determinants of money supply and money demand.

First, consider money supply. In the preceding chapter we discussed how the central bank, together with the banking system, determines the supply of money. When the central bank sells bonds in open-market operations, it receives money in exchange and contracts the money supply. When the central bank buys government bonds, it pays out money and expands the money supply. In addition, if any of this money

is deposited in banks which then hold it as reserves, the money multiplier swings into action, and these open-market operations can have an even greater effect on the money supply. For our purposes in this chapter, we ignore the complications introduced by the banking system and simply take the quantity of money supplied as a policy variable that the central bank controls.

Now consider money demand. Most fundamentally, the demand for money reflects how much wealth people want to hold in liquid form. Many factors influence the quantity of money demanded. The amount of currency that people hold in their wallets, for instance, depends on how much they rely on credit cards and on whether an automatic cash dispenser is easy to find. And, as we will emphasize later in the book, the quantity of money demanded depends on the interest rate that a person could earn by using the money to buy an interest-bearing bond rather than leaving it in their pocket or low-interest bank account.

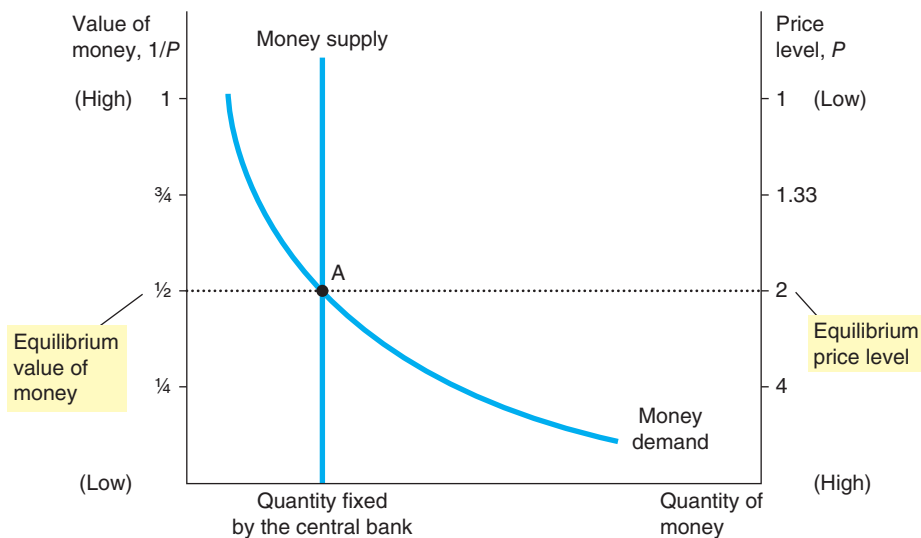
Although many variables affect the demand for money, one variable stands out in importance: the average level of prices in the economy. People hold money because it is the medium of exchange. Unlike other assets, such as bonds or stocks, people can use money to buy the goods and services on their shopping lists. How much money they choose to hold for this purpose depends on the prices of those goods and services. The higher prices are, the more money the typical transaction requires, and the more money people will choose to hold in their pockets and bank accounts. That is, a higher price level (a lower value of money) increases the quantity of money demanded.

What ensures that the quantity of money the central bank supplies balances the quantity of money people demand? The answer, it turns out, depends on the time horizon being considered. Later in this book we will examine the short-run answer, and we will see that interest rates play a key role. In the long run, however, the answer is different and much simpler. *In the long run, the overall level of prices adjusts to the level at which the demand for money equals the supply.* Figure 27.1 illustrates this idea. The horizontal axis of this graph shows the quantity of money. The left-hand vertical axis shows the value of money $1/P$, and the right-hand vertical axis shows the price level P . Notice that the price level axis on the right is inverted: a low price level is shown near the top of this axis, and a high price level is shown near the bottom. This inverted axis illustrates that when the value of money is high (as shown near the top of the left axis), the price level is low (as shown near the top of the right axis).

FIGURE 27.1

How the Supply and Demand for Money Determine the Equilibrium Price Level

The horizontal axis shows the quantity of money. The left vertical axis shows the value of money, and the right vertical axis shows the price level. The supply curve for money is vertical because the quantity of money supplied is fixed by the central bank. The demand curve for money is downward sloping because people want to hold a larger quantity of money when each euro buys less. At the equilibrium, point A, the value of money (on the left axis) and the price level (on the right axis) have adjusted to bring the quantity of money supplied and the quantity of money demanded into balance.



The two curves in this figure are the supply and demand curves for money. The supply curve is vertical because the central bank has fixed the quantity of money available. The demand curve for money is downwards sloping, indicating that when the value of money is low (and the price level is high), people demand a larger quantity of it to buy goods and services. At the equilibrium, shown in the figure as point A, the quantity of money demanded balances the quantity of money supplied. This equilibrium of money supply and money demand determines the value of money and the price level.

To summarize, if the price level is above the equilibrium level, people will want to hold more money than the central bank has created, so the price level must fall to balance supply and demand. If the price level is below the equilibrium level, people will want to hold less money than the central bank has created, and the price level must rise to balance supply and demand. At the equilibrium price level, the quantity of money that people want to hold exactly balances the quantity of money supplied by the central bank.

The Effects of a Monetary Injection

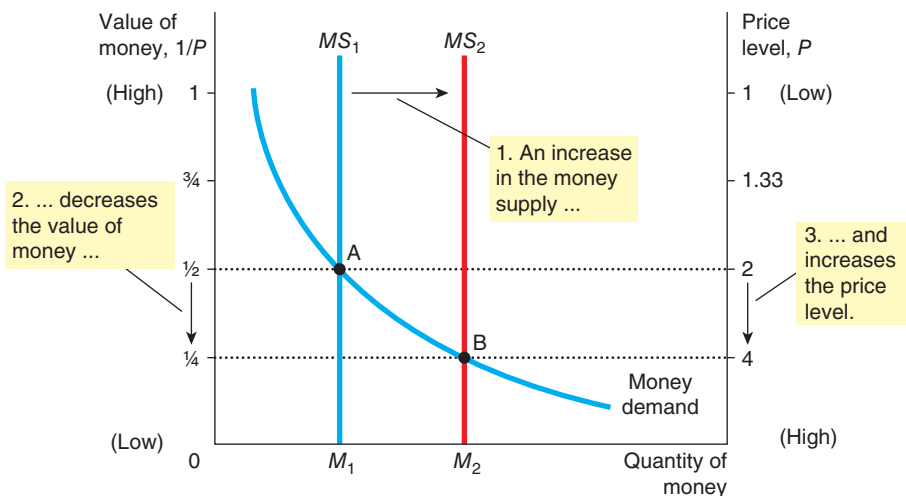
Let's now consider the effects of a change in monetary policy. To do so, imagine that the economy is in equilibrium and then, suddenly, the central bank doubles the supply of money by printing large amounts of money and dropping it around the country from helicopters (this analogy was first drawn by Milton Friedman and later echoed by the Federal Reserve chair, at the time of writing Ben Bernanke, which earned him the nickname 'Helicopter Ben'). Or, less dramatically and more realistically, the central bank could inject money into the economy by buying some government bonds from the public in open-market operations. What happens after such a monetary injection? How does the new equilibrium compare to the old one?

Figure 27.2 shows what happens. The monetary injection shifts the supply curve to the right from MS_1 to MS_2 , and the equilibrium moves from point A to point B. As a result, the value of money (shown on the left axis) decreases from $\frac{1}{2}$ to $\frac{1}{4}$, and the equilibrium price level (shown on the right axis) increases from 2 to 4. In other words, when an increase in the money supply makes euros more plentiful, the result is an increase in the price level that makes each euro less valuable.

FIGURE 27.2

An Increase in the Money Supply

When the central bank increases the supply of money, the money supply curve shifts from MS_1 to MS_2 . The value of money (on the left axis) and the price level (on the right axis) adjust to bring supply and demand back into balance. The equilibrium moves from point A to point B. Thus, when an increase in the money supply makes euros more plentiful, the price level increases, making each euro less valuable.



This explanation of how the price level is determined and why it might change over time is called the **quantity theory of money**. According to the quantity theory, the quantity of money available in the economy determines the value of money, and growth in the quantity of money is the primary cause of inflation. As Milton Friedman once put it, ‘Inflation is always and everywhere a monetary phenomenon.’

quantity theory of money a theory asserting that the quantity of money available determines the price level and that the growth rate in the quantity of money available determines the inflation rate

A Brief Look at the Adjustment Process

So far we have compared the old equilibrium and the new equilibrium after an injection of money. How does the economy get from the old to the new equilibrium? A complete answer to this question requires an understanding of short-run fluctuations in the economy, which we examine later in this book. Yet, even now, it is instructive to consider briefly the adjustment process that occurs after a change in money supply.

The immediate effect of a monetary injection is to create an excess supply of money. Before the injection, the economy was in equilibrium (point A in Figure 27.2). At the prevailing price level, people had exactly as much money as they wanted. But after the helicopters drop the new money and people pick it up off the streets, people have more euros in their pockets than they need to buy goods and services. At the prevailing price level, the quantity of money supplied now exceeds the quantity demanded.

People try to get rid of this excess supply of money in various ways. They might buy goods and services with their excess holdings of money. Or they might use this excess money to make loans to others by buying bonds or by depositing the money in a bank savings account (remember the relationship between saving and investment outlined in Chapter 24). These loans allow other people to buy goods and services. In either case, the injection of money increases the demand for goods and services.

The economy’s ability to supply goods and services, however, has not changed. As we saw in Chapter 22, the economy’s output of goods and services is determined by the available labour, physical capital, human capital, natural resources and technological knowledge. None of these is altered by the injection of money.

Thus, the greater demand for goods and services causes the prices of goods and services to increase. The increase in the price level, in turn, increases the quantity of money demanded because people are using more euros for every transaction. Eventually, the economy reaches a new equilibrium (point B in Figure 27.2) at which the quantity of money demanded again equals the quantity of money supplied. In this way, the overall price level for goods and services adjusts to bring money supply and money demand into balance.

The Classical Dichotomy and Monetary Neutrality

We have seen how changes in the money supply lead to changes in the average level of prices of goods and services. How do these monetary changes affect other important macroeconomic variables, such as production, employment, real wages and real interest rates? This question has long intrigued economists. Indeed, the great classical economist and philosopher David Hume wrote about it in the 18th century. The answer we give today owes much to Hume’s analysis.

Hume and his contemporaries suggested that all economic variables should be divided into two groups. The first group consists of **nominal variables** – variables measured in monetary units. The second group consists of **real variables** – variables measured in physical units. For example, the income of corn farmers is a nominal variable because it is measured in euros, whereas the quantity of corn they produce is a real variable because it is measured in kilos. Similarly, nominal GDP is a nominal variable because it measures the euro value of the economy’s output of goods and services; real GDP is a real variable because it measures the total quantity of goods and services produced and is not influenced by the current prices of those goods and services. This separation of variables into these groups is now called the **classical dichotomy**. (A *dichotomy* is a division into two groups, and *classical* refers to the earlier economic thinkers or classical economists.)

nominal variables variables measured in monetary units

real variables variables measured in physical units

classical dichotomy the theoretical separation of nominal and real variables

Application of the classical dichotomy is somewhat tricky when we turn to prices. Prices in the economy are normally quoted in terms of money and, therefore, are nominal variables. For instance, when we say that the price of corn is €2 a kilo or that the price of wheat is €1 a kilo, both prices are nominal variables. But what about a *relative* price – the price of one thing compared to another? In our example, we could say that the price of a kilo of corn is two kilos of wheat. Notice that this relative price is no longer measured in terms of money. When comparing the prices of any two goods, the euro signs cancel, and the resulting number is measured in physical units. The lesson is that money prices (e.g. in pounds, euros or dollars) are nominal variables, whereas relative prices are real variables.

This lesson has several important applications. Think about wages. Imagine that a consumer only ever buys bananas and that the price of a banana is €2. If the consumer's wage is €10 per hour then the consumer can buy five bananas with their wage. Ten euro per hour is the nominal wage rate measured in money terms. The **real wages** rate is given by the ratio of the wage rate to the price of bananas $\frac{W}{P}$. In this example the real wage rate is the number of bananas the consumer can buy with their wage, $\frac{10}{2} = 5$ bananas per hour. In order to be able to afford five bananas the consumer has to work for an hour. If the wage rate and prices change then the real wage rate gives a more accurate reflection of how the consumer has been affected. If the wage rate increases to €12 per hour and the price of bananas rises to €3 is the consumer better off or worse off? The real wage is now $\frac{12}{3} = 4$ bananas per hour. We can say that the worker is now worse off since they can now only buy four bananas for every hour they work rather than five. The real wage (the money wage adjusted for inflation) is a real variable because it measures the rate at which the economy exchanges goods and services for each unit of labour. Similarly, the real interest rate (the nominal interest rate adjusted for inflation) is a real variable because it measures the rate at which the economy exchanges goods and services produced today for goods and services produced in the future.

real wages the money wage adjusted for inflation measured by the ratio of the wage rate to price $\frac{W}{P}$

Why bother separating variables into these two groups? Hume suggested that the classical dichotomy is useful in analysing the economy because different forces influence real and nominal variables. In particular, he argued, nominal variables are heavily influenced by developments in the economy's monetary system, whereas the monetary system is largely irrelevant for understanding the determinants of important real variables.

Notice that Hume's idea was implicit in our earlier discussions of the real economy in the long run. In previous chapters we examined how real GDP, saving, investment, real interest rates and unemployment are determined without any mention of the existence of money. As explained in that analysis, the economy's production of goods and services depends on productivity and factor supplies, the real interest rate adjusts to balance the supply and demand for loanable funds, the real wage adjusts to balance the supply and demand for labour, and unemployment results when the real wage is for some reason kept above its equilibrium level. These important conclusions have nothing to do with the quantity of money supplied.

Changes in the supply of money, according to Hume, affect nominal variables but not real variables. When the central bank doubles the money supply, the price level doubles, the euro wage doubles, and all other euro values double. Real variables, such as production, employment, real wages and real interest rates, are unchanged. This irrelevance of monetary changes for real variables is called **monetary neutrality**.

monetary neutrality the proposition that changes in the money supply do not affect real variables

An analogy sheds light on the meaning of monetary neutrality. Recall that, as the unit of account, money is the yardstick we use to measure economic transactions. When a central bank doubles the money supply, all prices double, and the value of the unit of account falls by half. A similar change would occur if a European Union directive reduced the definition of the metre from 100 to 50 centimetres: as a result of the new unit of measurement, all *measured* distances (nominal variables) would double, but the *actual* distances (real variables) would remain the same. The euro, like the metre, is merely a unit of measurement, so a change in its value should not have important real effects.

Is this conclusion of monetary neutrality a realistic description of the world in which we live? The answer is, not completely. A change in the length of the metre from 100 to 50 centimetres would not matter much in the long run, but in the short run it would certainly lead to confusion and various mistakes. Similarly, most economists today believe that over short periods of time – within the span of a year or two – there is reason to think that monetary changes do have important effects on real variables. Hume himself also doubted that monetary neutrality would apply in the short run. (We will turn to the study of short-run non-neutrality later in the book, and this topic will shed light on the reasons why central banks change the supply of money over time.)

Most economists today accept Hume's conclusion as a description of the economy in the long run. Over the course of a decade, for instance, monetary changes have important effects on nominal variables (such as the price level) but only negligible effects on real variables (such as real GDP). When studying long-run changes in the economy, the neutrality of money offers a good description of how the world works.

Velocity and the Quantity Equation

We can obtain another perspective on the quantity theory of money by considering the following question: how many times per year is the typical €1 coin used to pay for a newly produced good or service? The answer to this question is given by a variable called the **velocity of money**. In physics, the term *velocity* refers to the speed at which an object travels. In economics, the velocity of money refers to the speed at which money changes hands as it moves around the economy.

velocity of money the rate at which money changes hands

To calculate the velocity of money, we divide the nominal value of output (nominal GDP) by the quantity of money. If P is the price level (the GDP deflator), Y the quantity of output (real GDP) and M the quantity of money, then velocity is:

$$V = \frac{(P \times Y)}{M}$$

To see why this makes sense, imagine a simple economy that produces only pizza. Suppose that the economy produces 100 pizzas in a year, that a pizza sells for €10, and that the quantity of money in the economy is €50, made up of fifty €1 coins. Then the velocity of money is:

$$\begin{aligned} V &= \frac{(\text{€}10 \times 100)}{\text{€}50} \\ &= 20 \end{aligned}$$

In this economy, people spend a total of €1,000 per year on pizza. For this €1,000 of spending to take place with only €50 of money, each euro coin must be spent (i.e. change hands) on average 20 times per year.

With slight algebraic rearrangement, this equation can be rewritten as:

$$M \times V = P \times Y$$

This equation states that the quantity of money (M) times the velocity of money (V) equals the price of output (P) times the amount of output (Y). It is called the **quantity equation** because it relates the quantity of money (M) to the nominal value of output ($P \times Y$). The quantity equation (which is an identity or a truism) shows that an increase in the quantity of money in an economy must be reflected in one of the other three variables: the price level must rise, the quantity of output must rise or the velocity of money must fall.

quantity equation the equation $M \times V = P \times Y$, which relates the quantity of money, the velocity of money, and the currency value of the economy's output of goods and services

We now have all the elements necessary to explain the equilibrium price level and inflation rate. Here they are:

1. Assume that the velocity of money is relatively stable over time.
2. Because velocity is stable, when the central bank changes the quantity of money (M), it causes proportionate changes in the nominal value of output ($P \times Y$).
3. The economy's output of goods and services (Y) is primarily determined by factor supplies (labour, physical capital, human capital and natural resources) and the available production technology. In particular, because money is neutral, money does not affect output.
4. With output (Y) determined by factor supplies and technology, when the central bank alters the money supply (M) and induces proportional changes in the nominal value of output ($P \times Y$), these changes are reflected in changes in the price level (P).
5. Therefore, when the central bank increases the money supply rapidly, the result is a high rate of inflation.

These five steps are the essence of the quantity theory of money.

CASE STUDY The Velocity of Circulation

The discussion of the quantity theory of money above makes an important assumption – that the velocity of circulation of money (V) is stable over time. How accurate is such an assumption? One part of the answer might be to consider what is meant by 'stable' and over what time period we are looking. Stable does not mean constant but what degree of volatility is acceptable to describe V as 'stable'?

Using the quantity theory equation, it can be seen that if V is not assumed constant then changes in the money stock have different implied effects on inflation. If we observe that the money stock has increased by (say) 10 per cent but the price level remained unchanged then either the velocity of circulation must have slowed down or output might have fallen. The more stable V is, the more powerful is the quantity theory in helping to monitor the behaviour of the economy and the role of monetary policy.

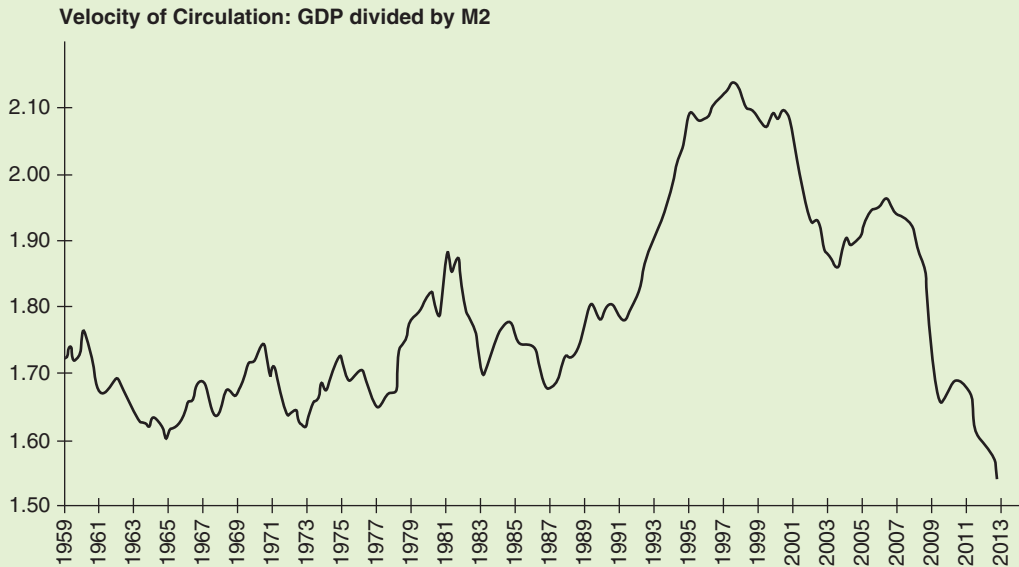
Figure 27.3 shows a measure of V in the United States taken by taking US GDP and dividing by the Federal Reserve's M2 measure of the money supply from 1959 to 2013. The figure shows V fluctuating between a corridor of 1.6 and 1.9 between 1959 and 1993 but appearing to be a lot more volatile in the subsequent 20 years. Overall, V appears to have varied between a corridor of 1.5 and 2.10 over the time period shown – a variation of around 0.6 over the period. Can this degree of fluctuation be defined as 'stable'?

Is the argument about the stability of V actually relevant? Whilst many economists do place some value in the quantity theory, there are some who cast doubt on it and they do so because of the velocity of circulation which they believe is a fallacy. In Chapter 26 we noted that the Austrian school of economists had a definition of the true money supply. It is the Austrian school who cast doubt on the validity of V .

Their argument is as follows. When individuals make market transactions they pay through the value of the goods and services they produce themselves. A university lecturer, in reality, pays for the food she consumes with the value of the lecture services she produces. Money is merely the means to facilitate

FIGURE 27.3**The Velocity of Circulation, 1959–2013**

The velocity of circulation in this figure is measured by taking US GDP from 1959–2013 and dividing it by the US Federal Reserve's M2 measure of the money supply.



the exchange – the medium of exchange. The number of times a €10 note changes hands has nothing to do with the lecturer's ability to be able to fund her desired purchase of food; it is the lecturer's ability to provide teaching services which determines her ability to buy food. What happens, so the argument goes, is that the lecturer is exchanging her services for money which then enables her to make transactions. The food retailer is exchanging the services he provides for money and using that money to buy the goods and services he needs. In making these transactions both parties are assigning a value to the transactions they make in relation to their own particular well-being. The Austrian school argues that it is individuals' actions that determine the prices of goods and services and not the speed at which money changes hands. V cannot, therefore, say anything about the average price level or the average purchasing power of money. V is simply the value of transactions, the price multiplied by output, divided by the money stock $V = P \left(\frac{Y}{M} \right)$ and as such is a truism which is not helpful and V cannot be defined independently of the rest of the equation – it needs P , Y and M to give it any meaning and as such V in itself cannot be a causal factor of anything. This implies that if M is rising then if V was declining it cannot offset the growth in the price level.

The Austrian school link V to the demand for money. If V is unstable does that also mean that the demand for money is unstable? The Austrian school see this statement as rather absurd – if individuals change their decisions to hold more or less money over time as a result of changes in their personal circumstances, why would this be an issue? Their view would be that over time the demand for money changes – in the same way that over time the demand for any good or service changes. It would be absurd to expect the demand for fish to be stable over time so why should we treat the demand for money in any different way?

This case study reflects the fact that different economists look at theories in different ways. The Austrian school does not deny a link between the money supply and prices but questions the quantity theory (or the equation of exchange) as being an accurate way of looking at the economy. In doing so, they argue, false conclusions may be drawn.

The Inflation Tax

If inflation is so easy to explain, why do countries experience hyperinflation? That is, why do the central banks of these countries choose to print so much money that its value is certain to fall rapidly over time?

The answer is that the governments of these countries are using money creation as a way to pay for their spending. When the government wants to build roads, pay salaries to police officers, or give transfer payments to the poor or elderly, it first has to raise the necessary funds. Normally, the government does this by levying taxes, such as income and sales taxes, and by borrowing from the public by selling government bonds. Yet the government can also pay for spending by simply printing the money it needs.

When the government raises revenue by printing money, it is said to levy an **inflation tax**. The inflation tax is not exactly like other taxes, however, because no one receives a bill from the government for this tax. Instead, the inflation tax is more subtle. When the government prints money, the price level rises, and the euros in your pocket are less valuable. Thus, the inflation tax is like a tax on everyone who holds money. It is even, roughly speaking, a progressive tax since the richer you are, the more money you are likely to hold and therefore the more the inflation tax will affect you.

inflation tax the revenue the government raises by creating money

The importance of the inflation tax varies from country to country and over time. In the early 21st century, most advanced economies have been enjoying very low rates of inflation, so that the inflation tax has been a trivial source of revenue for the government, amounting to only a few per cent of government revenue. In periods of hyperinflation in Germany of the 1920s, in various Latin American countries during the 1970s and 1980s and in Yugoslavia and Zimbabwe more recently, the inflation tax would have been quite considerable.

Nearly all hyperinflations follow the same pattern: the government has high spending, limited ability to borrow, inadequate tax revenue (perhaps because the level of income in the economy is low, or there is widespread tax evasion or a poorly developed tax system, or a combination of all these factors) and, as a result, it turns to the printing press to pay for its spending. The massive increases in the quantity of money lead to massive inflation. The inflation ends when the government institutes fiscal reforms – such as cuts in government spending – that eliminate the need for the inflation tax.

As the economist John Maynard Keynes once pointed out, the temptation to just print money to pay for government spending through the inflation tax may be hard for a government to resist in such circumstances: 'The burden of the tax is well spread, cannot be evaded, costs nothing to collect, and falls, in a rough sort of way, in proportion to the wealth of the victim. No wonder its superficial advantages have attracted Ministers of Finance.' Of course, Keynes also recognized the great damage that high inflation can do to an economy, which is why he referred to the advantages of an inflation tax as 'superficial'.

The Fisher Effect

According to the principle of monetary neutrality, an increase in the rate of money growth raises the rate of inflation but does not affect any real variable. An important application of this principle concerns the effect of money on interest rates. Interest rates are important variables for macroeconomists to understand because they link the economy of the present and the economy of the future through their effects on saving and investment.

To understand the relationship between money, inflation and interest rates, recall the distinction between the nominal interest rate and the real interest rate.

The *nominal interest rate* is the interest rate you hear about at your bank. If you have a savings account, for instance, the nominal interest rate tells you how fast the number of euros in your account will rise over time. The *real interest rate* corrects the nominal interest rate for the effect of inflation in order to tell you how fast the purchasing power of your savings account will rise over time. The real interest rate is the nominal interest rate minus the inflation rate:

$$\text{Real interest rate} = \text{Nominal interest rate} - \text{Inflation rate}$$

For example, if the bank posts a nominal interest rate of 7 per cent per year and the inflation rate is 3 per cent per year, then the real value of the deposits grows by 4 per cent per year.

We can rewrite this equation to show that the nominal interest rate is the sum of the real interest rate and the inflation rate:

$$\text{Nominal interest rate} = \text{Real interest rate} + \text{Inflation rate}$$

This way of looking at the nominal interest rate is useful because different economic forces determine each of the two terms on the right-hand side of this equation. As we discussed earlier in the book, the supply and demand for loanable funds determine the real interest rate. And, according to the quantity theory of money, growth in the money supply determines the inflation rate.

Let's now consider how the growth in the money supply affects interest rates. In the long run over which money is neutral, a change in money growth should not affect the real interest rate. The real interest rate is, after all, a real variable. For the real interest rate not to be affected, the nominal interest rate must adjust one-for-one to changes in the inflation rate. Thus, when the central bank increases the rate of money growth, the result is both a higher inflation rate and a higher nominal interest rate. This adjustment of the nominal interest rate to the inflation rate is called the **Fisher effect**, after the American economist Irving Fisher (1867–1947), who first studied it.

Fisher effect the one-for-one adjustment of the nominal interest rate to the inflation rate

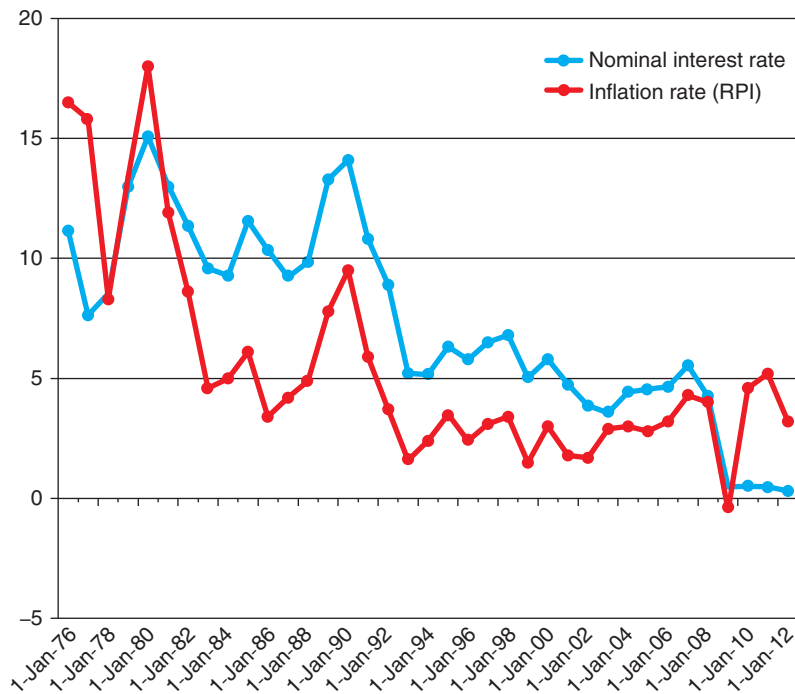
Keep in mind that our analysis of the Fisher effect has maintained a long-run perspective. The Fisher effect does not hold in the short run to the extent that inflation is unanticipated. A nominal interest rate is a payment on a loan, and it is typically set when the loan is first made. If inflation catches the borrower and lender by surprise, the nominal interest rate they set will fail to reflect the rise in prices. To be precise, the Fisher effect states that the nominal interest rate adjusts to expected inflation. Expected inflation moves with actual inflation in the long run but not necessarily in the short run.

The Fisher effect is crucial for understanding changes over time in the nominal interest rate. Figure 27.4 shows the nominal interest rate and the inflation rate in the UK economy since 1976. The close association between these two variables is clear. The nominal interest rate tends to rise when inflation rises and fall when inflation falls. This is true both in high-inflation periods, such as during the late 1970s and the late 1980s, as well as during low-inflation periods, such as the period between 1992 and 2002.

SELF TEST The government of a country increases the growth rate of the money supply from 5 per cent per year to 50 per cent per year. ● What happens to prices? ● What happens to nominal interest rates? ● Why might the government be doing this?

FIGURE 27.4**The UK Nominal Interest Rate and the Inflation Rate**

This figure uses annual data since 1976 to show the nominal interest rate on three-month UK Treasury bills (measured as the annual average rate of discount) and the inflation rate (as measured by the retail prices index). The close association between these two variables is evidence for the Fisher effect: when the inflation rate rises, so does the nominal interest rate.



THE COSTS OF INFLATION

In the mid-1970s, when the UK inflation rate reached a peak of about 24 per cent a year, inflation dominated debates over economic policy. And, even though inflation was low during the 1990s, it remained a closely watched macroeconomic variable.

Inflation is closely watched and widely discussed because it is thought to be a serious economic problem. But is that true? And if so, why?

A Fall in Purchasing Power? The Inflation Fallacy

If you ask the typical person why inflation is bad, he will tell you that the answer is obvious: inflation robs him of the purchasing power of his hard-earned money. When prices rise, each euro of income buys fewer goods and services. Thus, it might seem that inflation directly lowers living standards.

Yet further thought reveals a fallacy in this answer. When prices rise, buyers of goods and services pay more for what they buy. At the same time, however, sellers of goods and services get more for what they sell. Because most people earn their incomes by selling their services, such as their labour, inflation in incomes goes hand in hand with inflation in prices. *Thus, inflation does not in itself reduce people's real purchasing power.*

People believe the inflation fallacy because they do not appreciate the principle of monetary neutrality. A worker who receives an annual rise of 10 per cent in her salary tends to view that rise as a reward for her own talent and effort. When an inflation rate of 6 per cent reduces the real value of that pay rise to only 4 per cent, the worker might feel that she has been cheated of what is rightfully her due. In fact, as we discussed in Chapter 22, real incomes are determined by real variables, such as physical capital, human

capital, natural resources and the available production technology. Nominal incomes are determined by those factors and the overall price level. If the central bank were to succeed in lowering the inflation rate from 6 per cent to zero, our worker's annual rise would fall from 10 per cent to 4 per cent. She might feel less robbed by inflation, but her real income would not rise more quickly.

If nominal incomes tend to keep pace with rising prices, why then is inflation a problem? It turns out that there is no single answer to this question. Instead, economists have identified several costs of inflation. Each of these costs shows some way in which persistent growth in the money supply does, in fact, have some effect on real variables.

Shoeleather Costs

As we have discussed, inflation is like a tax on the holders of money. The tax itself is not a cost to society: it is only a transfer of resources from households to the government. Yet most taxes give people an incentive to alter their behaviour to avoid paying the tax, and this distortion of incentives causes deadweight losses for society as a whole. Like other taxes, the inflation tax also causes deadweight losses because people waste scarce resources trying to avoid it.

How can a person avoid paying the inflation tax? Because inflation erodes the real value of the money in your pocket, you can avoid the inflation tax by holding less money. One way to do this is to go to the bank more often. For example, rather than withdrawing €400 every four weeks, you might withdraw €100 once a week. By making more frequent trips to the bank, you can keep more of your wealth in your interest-bearing savings account and less in your pocket, where inflation erodes its value.

The cost of reducing your money holdings is called the **shoeleather cost** of inflation because making more frequent trips to the bank causes your shoes to wear out more quickly. Of course, this term is not to be taken literally: the actual cost of reducing your money holdings is not the wear and tear on your shoes but the time and convenience you must sacrifice to keep less money on hand than you would if there were no inflation – it is in effect a **transaction cost**, the opportunity cost of carrying out the trips to the bank.

shoeleather cost the resources wasted when inflation encourages people to reduce their money holdings
transaction cost the opportunity cost of carrying out a transaction in any market

Shoeleather costs of inflation may seem trivial. And, in fact, they are in countries experiencing only moderate inflation. But this cost is magnified in countries experiencing hyperinflation. In conditions of hyperinflation, shoeleather costs can be substantial. With a high inflation rate, individuals do not have the luxury of holding the local money as a store of value. Instead, they are forced to convert the domestic currency quickly into goods or into another currency, often the US dollar, which offer a more stable store of value. The time and effort that individuals expend to reduce their money holdings are a waste of resources. If monetary authorities pursued a low-inflation policy, individuals would be happy to hold the domestic currency, and they could put time and effort to more productive use.

Menu Costs

Most firms do not change the prices of their products every day. Instead, firms often announce prices and leave them unchanged for weeks, months or even years.

Firms change prices infrequently because there are costs involved in changing prices. Costs of price adjustment are called **menu costs**, a term derived from a restaurant's cost of printing a new menu. Menu costs include the cost of deciding on new prices, the cost of printing new price lists and catalogues, the cost of sending these new price lists and catalogues to dealers and customers, the cost of advertising the new prices, and even the cost of dealing with customer annoyance over price changes.

menu costs the costs of changing prices

Inflation increases the menu costs that firms must bear. In an economy with low inflation of just a few percentage points a year, annual price adjustment is an appropriate business strategy for many firms. But when high inflation makes firms' costs rise rapidly, annual price adjustment is impractical. During hyperinflations, for example, firms must change their prices daily or even more often just to keep up with all the other prices in the economy. At the height of the German hyperinflation in the 1920s, people eating in restaurants would sometimes insist on paying the bill at the beginning of the meal rather than at the end, because the price of the food would rise while they were eating!

Relative Price Variability and the Misallocation of Resources

Suppose that a fast food restaurant prints a new menu with new prices every January and then leaves its prices unchanged for the rest of the year. If there is no inflation, the restaurant's relative prices – the prices of its meals compared with other prices in the economy – would be constant over the course of the year. By contrast, if the inflation rate is 12 per cent per year, the restaurant's relative prices will automatically fall by 1 per cent each month. The restaurant's relative prices (that is, its prices compared with others in the economy) will be high in the early months of the year, just after it has printed a new menu, and low in the later months. And the higher the inflation rate, the greater is this automatic variability. Thus, because prices change only once in a while, inflation causes relative prices to vary more than they otherwise would.

Why does this matter? The reason is that market economies rely on relative prices to allocate scarce resources. Consumers decide what to buy by comparing the quality and prices of various goods and services. Through these decisions, they determine how the scarce factors of production are allocated among industries and firms. When inflation distorts relative prices, consumer decisions are distorted, and markets are less able to allocate resources to their best use.

Inflation-Induced Tax Distortions

Almost all taxes distort incentives, cause people to alter their behaviour, and lead to a less efficient allocation of the economy's resources. Many taxes, however, become even more problematic in the presence of inflation. The reason is that politicians often fail to take inflation into account when writing the tax laws. Economists who have studied the tax system conclude that inflation tends to raise the tax burden on income earned from savings.

One example of how inflation discourages saving is the tax treatment of *capital gains* – the profits made by selling an asset for more than its purchase price. Suppose that in 2005 you used some of your savings to buy stock in Apple for €10 and that in 2015 you sold the stock for €50. According to the tax system, you have earned a capital gain of €40, which you must include in your income when computing how much income tax you owe. But suppose the overall price level doubled from 2005 to 2015. In this case, the €10 you invested in 2005 is equivalent (in terms of purchasing power) to €20 in 2015. When you sell your Apple shares for €50, you have a real gain (an increase in purchasing power) of only €30. The tax system, however, does not take account of inflation and assesses you a tax on a gain of €40. Thus, inflation exaggerates the size of capital gains and inadvertently increases the tax burden on this type of income.

Another example is the tax treatment of interest income. The income tax treats the *nominal* interest earned on savings as income, even though part of the nominal interest rate merely compensates for inflation. To see the effects of this policy, consider the numerical example in Table 27.1. The table compares two economies, both of which tax interest income at a rate of 25 per cent. In economy A, inflation is zero, and the nominal and real interest rates are both 4 per cent. In this case, the 25 per cent tax on interest income reduces the real interest rate from 4 per cent to 3 per cent. In economy B, the real interest rate is again 4 per cent, but the inflation rate is 8 per cent. As a result of the Fisher effect, the nominal interest rate is 12 per cent. Because the income tax treats this entire 12 per cent interest as income, the government takes 25 per cent of it, leaving an after-tax nominal interest rate of only 9 per cent and an after-tax real interest rate of only 1 per cent. In this case, the 25 per cent tax on interest income reduces the real interest rate from 4 per cent to 1 per cent. Because the after-tax real interest rate provides the incentive to save, saving is much less attractive in the economy with inflation (economy B) than in the economy with stable prices (economy A).

TABLE 27.1 How Inflation Raises the Tax Burden on Saving

In the presence of zero inflation, a 25 per cent tax on interest income reduces the real interest rate from 4 per cent to 3 per cent. In the presence of 8 per cent inflation, the same tax reduces the real interest rate from 4 per cent to 1 per cent.

| | Economy A (price stability) | Economy B (inflation) |
|---|--|----------------------------------|
| Real interest rate | 4% | 4% |
| Inflation rate | 0 | 8 |
| Nominal interest rate (real interest rate + inflation rate) | 4 | 12 |
| Reduced interest due to 25 per cent tax (.25 × nominal interest rate) | 1 | 3 |
| After-tax nominal interest rate (.75 × nominal interest rate) | 3 | 9 |
| After-tax real interest rate (after-tax nominal interest rate – inflation rate) | 3 | 1 |

The taxes on nominal capital gains and on nominal interest income are two examples of how the tax system interacts with inflation. There are many others. Because of these inflation-induced tax changes, higher inflation tends to discourage people from saving. Recall that the economy's saving provides the resources for investment, which in turn is a key ingredient to long-run economic growth. Thus, when inflation raises the tax burden on saving, it tends to depress the economy's long-run growth rate. There is, however, no consensus among economists about the size of this effect.

One solution to this problem, other than eliminating inflation, is to index the tax system. That is, the tax laws could be rewritten to take account of the effects of inflation. In the case of capital gains, for example, the tax code could adjust the purchase price using a price index and assess the tax only on the real gain. In the case of interest income, the government could tax only real interest income by excluding that portion of the interest income that merely compensates for inflation.

In an ideal world, tax laws would be written so that inflation would not alter anyone's real tax liability. In the world in which we live, however, tax laws are far from perfect. More complete indexation would probably be desirable, but it would further complicate a tax system that many people already consider too complex.

Confusion and Inconvenience

Imagine that we took a poll and asked people the following question: 'This year the metre is 100 centimetres. How long do you think it should be next year?' Assuming we could get people to take us seriously, they would tell us that the metre should stay the same length – 100 centimetres. Anything else would just complicate life needlessly.

What does this finding have to do with inflation? Recall that money, as the economy's unit of account, is what we use to quote prices and record debts. In other words, money is the yardstick with which we measure economic transactions. The job of the central bank is a little like the job of the government department that deals with weights and measurements, i.e. to ensure the reliability of a commonly used unit of measurement. When the central bank increases the money supply and creates inflation, it erodes the real value of the unit of account.

It is difficult to judge the costs of the confusion and inconvenience that arise from inflation. Earlier we discussed how the tax system incorrectly measures real incomes in the presence of inflation. Similarly, accountants incorrectly measure firms' earnings when prices are rising over time. Because inflation causes money at different times to have different real values, computing a firm's profit – the difference between its revenue and costs – is more complicated in an economy with inflation. Therefore, to some extent, inflation makes investors less able to sort out successful from unsuccessful firms, which in turn impedes financial markets in their role of allocating the economy's saving to alternative types of investment.

A Special Cost of Unexpected Inflation: Arbitrary Redistributions of Wealth

So far, the costs of inflation we have discussed occur even if inflation is steady and predictable. Inflation has an additional cost, however, when it comes as a surprise. Unexpected inflation redistributes wealth among the population in a way that has nothing to do with either merit or need. These redistributions occur because many loans in the economy are specified in terms of the unit of account – money.

Consider an example. Suppose that Lars, a student, takes out a €20,000 loan at a 7 per cent interest rate from Bigbank to attend university. In ten years the loan will have to be repaid. After his debt has compounded for ten years at 7 per cent, Lars will owe Bigbank €40,000 (rounded). The real value of this debt will depend on inflation over the decade. If Lars is lucky, the economy will have hyperinflation. In this case, wages and prices will rise so high that Lars will be able to pay the €40,000 debt out of pocket change. In contrast, if the economy goes through a major deflation, then wages and prices will fall, and Lars will find the €40,000 debt a greater burden than he anticipated.

This example shows that unexpected changes in prices redistribute wealth among debtors and creditors. A hyperinflation enriches Lars at the expense of Bigbank because it diminishes the real value of the debt; Lars can repay the loan in less valuable euros than he anticipated. Deflation enriches Bigbank at Lars' expense because it increases the real value of the debt; in this case, Lars has to repay the loan in more valuable euros than he anticipated. If inflation were predictable, then Bigbank and Lars could take inflation into account when setting the nominal interest rate (recall the Fisher effect). But if inflation is hard to predict, it imposes risk on Lars and Bigbank that both would prefer to avoid.

This cost of unexpected inflation is important to consider together with another fact: inflation is especially volatile and uncertain when the average rate of inflation is high. This is seen most simply by examining the experience of different countries. Countries with low average inflation, such as Germany, tend to have stable inflation. Countries with high average inflation, such as many countries in Latin America, tend also to have unstable inflation. There are no known examples of economies with high, stable inflation. This relationship between the level and volatility of inflation points to another cost of inflation. If a country pursues a high-inflation monetary policy, it will have to bear not only the costs of high expected inflation but also the arbitrary redistributions of wealth associated with unexpected inflation.

The Price of Ice Cream

To return to our starting point – an important discussion of the value of ice cream – how does the price of the 1930 choc ice compare to one you could buy today? Remember that in the 1930s you could have bought 120 choc ices for a pound. Earlier in the book we examined how economists measure the inflation rate as the percentage change in the CPI, the GDP deflator, or some other index of the overall price level. These price indices show that, over the past 70 years or so, UK prices have risen on average about 5 per cent per year. Accumulated over so many years, a 5 per cent annual inflation rate leads to something like a 30-fold increase in the price level, so that £1 in 1930s money is equivalent to about £30 in today's money. With a small amount of mental arithmetic, we can see that the price of a Lyons Maid choc ice in the 1930s would be about 25p in today's money. The retail price of about 70p for a Wall's 'Chunky' choc ice nowadays therefore represents a substantial real increase in the price of a chocolate-coated ice-cream bar, even allowing for the possibility of higher quality. Perhaps people have indeed come to enjoy ice cream more.

SELF TEST List and describe six costs of inflation.

Deflation

The majority of this chapter has focused on inflation – a period of generally rising prices. Economists generally see some inflation in the economy as being desirable (provided it is stable and manageable). The opposite of inflation is deflation where the price level actually falls – not to be confused with a slowdown in the rate of growth of prices. It might sound intuitive that falling prices would be a good thing but that is not necessarily the case.

The post-war period saw unprecedented economic growth in Japan as it grew to become the second biggest economy in the world. Its methods and models were copied and applied in most Western economies; they looked enviously at the consistently high levels of productivity and quality of goods coming out of Japan as its exports flooded the globe.

In the last 15 years, however, the Japanese economy has suffered a number of problems, not least a prolonged recession and deflation. Corporate profits have been low and wage levels falling as business tries to adjust to the new market situation. The result of this is that there is little growth, little incentive to invest, low levels of spending and an increased likelihood of rising unemployment.

Some inflation in any economy is seen as being desirable because it is a sign that demand is present, that there is a reason to produce and invest and that there will be reward to be gained from enterprise. For a number of years in the first decade of the 21st century, interest rates in Japan remained at zero in an effort to help stimulate consumer spending. As with any price in any market, the function of the interest rate is to act as a signal and as such provide incentives that allocate scarce resources. When interest rates are zero, there is very little incentive for institutions to lend money; for businesses it might be good news if money can be raised so cheaply for investment but then again, what is there to invest in? Part of the deflation problem in Japan is that consumers were not spending at the levels that would provide a reason for investing in new capacity. Why spend or invest today when you expect prices to be cheaper tomorrow? If consumers cut back spending then firms see order books shrinking and so cut back on production and lay off staff. Prices fall further as the economy falls into a deflationary spiral.

Deflation, therefore, can be as damaging as inflation and this is the reason why central banks are invariably charged with maintaining targets for inflation at or around 2 per cent. Some inflation is considered good for the economy, when prices rise rapidly it causes the problems outlined in this chapter but this does not mean that falling prices is any better.

CONCLUSION

This chapter discussed the causes and costs of inflation. The primary cause of inflation is simply growth in the quantity of money. When the central bank creates money in large quantities, the value of money falls quickly. To maintain stable prices, the central bank must maintain strict control over the money supply.

The costs of inflation are more subtle. They include shoeleather costs, menu costs, increased variability of relative prices, unintended changes in tax liabilities, confusion and inconvenience, and arbitrary redistributions of wealth. Are these costs, in total, large or small? All economists agree that they become huge during hyperinflation. But their size for moderate inflation – when prices rise by less than 10 per cent per year – is more open to debate.

Although this chapter presented many of the most important lessons about inflation, the discussion is incomplete. When the central bank reduces the rate of money growth, prices rise less rapidly, as the quantity theory suggests. Yet as the economy makes the transition to this lower inflation rate, the change in monetary policy will have disruptive effects on production and employment. That is, even though monetary policy is neutral in the long run, it has profound effects on real variables in the short run. Later in this book we will examine the reasons for short-run monetary non-neutrality in order to enhance our understanding of the causes and costs of inflation.

IN THE NEWS



The Money Supply and Inflation

It is often the case in economics that disagreements about policy options arise not because of whether a particular policy will work but what its consequences will be and the extent to which it will have a beneficial effect.

Helicopter Money or Quantitative Easing – Just a Question of Semantics?

The financial crisis precipitated a period of economic slowdown across the globe which has persisted for far longer than other similar recessionary periods. Calls for growth strategies from governments across Europe and the United States, in addition to the need to get government deficits under control have been made, but exactly what form will these growth strategies take? One policy adopted by central banks has been quantitative easing (QE), which has been popularly reported as ‘the government printing money’.

One concern that has been raised with such a policy is the possible long-term effect on prices – if billions of euros, pounds or dollars are being pumped into the economy then the money supply will surely rise and as the quantity theory suggests, will make it more likely that inflation will accelerate in the future. Some, therefore, question whether QE is actually any different to the idea of ‘helicopter money’.

Consider the following two scenarios. 1. Each individual in a country is given cash. Some of this cash will be spent and some of it will be saved. 2. If that same amount of cash was given to the government, it could use the money to provide tax cuts, be spent on new infrastructure projects or pay off its debts.

If a government chose a helicopter drop, it is in effect giving its citizens cash to spend. The government flies around the country scattering money out of its helicopter, the cash floats down onto the streets for anyone to pick up. Once people pick it up they then spend it but what they spend it on is not under the government’s control. What it does do, however, is to provide a much needed boost to demand in the economy (it increases consumption, C , and so boosts aggregate demand). Would such a boost to aggregate demand be sustainable and significant to help get the economy out of the doldrums? The danger is that if the country’s capacity to expand output to meet this increased demand does not expand also then aggregate demand will be greater than aggregate supply and inflation will be the result. The government cannot go round the towns and cities of the country asking people to give back the money it has scattered in order to control inflation. In terms of the quantity theory, if governments print more money through a helicopter drop it will lead to inflation.

Is QE any different to scattering money out of a helicopter? QE involves the central bank buying bonds and the money paid ends up in the financial system in the hope that banks and financial institutions will lend to businesses and consumers thus boosting demand. The end result is meant to be the same but

the long-run effects are different. Crucially, it is argued that QE does not lead to a long-run increase in the money supply.

However, part of QE involves the central bank buying government bonds and this has led to some concern that the difference between QE and a helicopter drop is actually very fine. In addition, there has been concern that QE is being used as a tool of monetary financing – in other words, the central bank giving money to government to finance spending, which is inflationary because it increases the money supply.

Why? Well, government bonds are issued to fund government deficits – the difference between tax receipts and government expenditure. The questions being asked, therefore, are whether QE is financing government spending which would be akin to overt monetary financing (OMF) and whether QE is the same as a helicopter drop. With governments struggling to manage deficits, it is possible to propose an argument along the lines of the following: if the central bank buys government bonds the interest rate on government bonds falls and as a result government borrowing costs are lower. In addition, if the central bank is buying government bonds, what happens to the interest on these bonds – should the interest be accounted for on the central bank’s accounts or on the government’s (since it has issued the bonds)? When central banks print

money to finance government spending this is called monetary financing. The European Central Bank (ECB) is forbidden under its charter to engage in monetary financing. Monetary financing is also frowned upon by many other countries but where is the tipping point? There is a fine line between central bank open market operations involving government bonds and monetary financing largely depending on how different transactions are defined.

In the UK, for example, the Treasury announced in November 2012 that it intended to credit accumulated interest payments on UK government bonds bought by the Bank of England under QE to its account. The total sum was estimated as being some £35 billion over two years – a not unhelpful amount when the government is trying to cut its budget deficit. Such an accounting procedure could be looked at as being pretty much the same as scenario 2 above. If the central bank did engage in monetary financing then the result would be a rise in the money supply and, given the quantity theory, a rise in inflation.

Some question whether quantitative easing is actually any different to the idea of 'helicopter money'.



The counter argument is that QE requires money to be 'printed' by central banks but not for handing to people to spend but for the purchase of securities in financial markets. In this way, what is happening is that assets are being transferred within the financial system – in some cases, private firms are transferring bonds to the central bank's accounts and receiving cash in return which is recorded in its account. In other cases, it is a transfer of assets between the government and the central bank – the central bank is acquiring government bonds and cash is indirectly going to the government. This might be seen as being pretty much the same as monetary financing except that in the long run the monetary base need not rise and as a result inflation may not result unlike overt monetary financing (helicopter drops).

In a helicopter drop, assume the government instructs the central bank to send every individual in the country a cheque for €5,000. This can be seen as being exactly the same as a one-off tax rebate by the government (or a temporary tax cut). How is this financed? The government issues

bonds equivalent to the total amount given out to its citizens and the central bank buys up those bonds and the interest on the bonds and the principal when the bond is redeemed on maturity all goes back the government.

QE might be seen as being something very similar but the government does have to pay interest on the bonds it issues and which the central bank buys. In the economic climate in which QE is being used, interest rates happen to

be very low but nevertheless, there is a cost to the government unlike a pure helicopter drop. Whilst interest rates remain low (close to zero) it does not look as though there is a great deal of difference between a helicopter drop and QE but at some point in the future, interest rates will rise when economies begin to grow more strongly again and then the government will incur a higher cost of issuing bonds.

With QE, the central bank may have bought bonds and expanding the money supply but at some point in the future, when the economic climate is much better, QE can be unwound in that the central bank can sell back the bonds. The central bank could begin to sell bonds if inflationary pressures started to build and in doing so could take money out of the economy thus reducing the money supply. Proponents of QE argue that unlike a helicopter drop, it is a more flexible policy which allows the central bank to monitor inflationary pressures and adjust its asset purchasing programme to ensure that if it is perceived that the programme is ceasing to add to aggregate demand and boosting economic activity but instead is starting to have an impact on prices, it can be reversed.

Whether QE has the same impact on aggregate demand in a period of economic downturn that a helicopter drop would have is another matter. Critics of QE have argued that the billions spent by central banks have not found their way through to the real economy (to people who spend more money on consumption) but instead have just been absorbed by the financial system to correct balance sheets. Ultimately the question is, do we want to boost the economy at a risk of much higher inflation in the future or do we take a more risk averse approach and try and boost the economy to a lesser extent but have more control in the long run over inflation?

(Continued)

Questions

- 1 What effect would an increase in the money supply of 10 per cent have on the price level? On what does your answer depend?
- 2 If helicopter money is dropped, why might the way in which it was handled by those who 'pick it up' be important in determining the extent to which such a policy would boost aggregate demand?
- 3 Why would government borrowing costs fall if the central bank buys government bonds as part of an asset purchasing programme (QE)?
- 4 Why do proponents of QE argue that the policy is not inflationary in the same way as a helicopter drop?
- 5 If you were a policymaker in a period of severe economic decline, would you favour a helicopter drop or a programme of QE? Justify your reasoning.

SUMMARY

- The overall level of prices in an economy adjusts to bring money supply and money demand into balance. When the central bank increases the supply of money, it causes the price level to rise. Persistent growth in the quantity of money supplied leads to continuing inflation.
- The principle of monetary neutrality asserts that changes in the quantity of money influence nominal variables but not real variables. Most economists believe that monetary neutrality approximately describes the behaviour of the economy in the long run.
- A government can pay for some of its spending simply by printing money. When countries rely heavily on this 'inflation tax', the result is hyperinflation.
- One application of the principle of monetary neutrality is the Fisher effect. According to the Fisher effect, when the inflation rate rises, the nominal interest rate rises by the same amount, so that the real interest rate remains the same.
- Many people think that inflation makes them poorer because it raises the cost of what they buy. This view is a fallacy, however, because inflation also raises nominal incomes.
- Economists have identified six costs of inflation: shoeleather costs associated with reduced money holdings; menu costs associated with more frequent adjustment of prices; increased variability of relative prices; unintended changes in tax liabilities due to non-indexation of the tax system; confusion and inconvenience resulting from a changing unit of account; and arbitrary redistributions of wealth between debtors and creditors. Many of these costs are large during hyperinflation, but the size of these costs for moderate inflation is less clear.
- Deflation occurs when prices actually fall and the effects on the economy and on incentives can be as damaging as when inflation is rising too quickly.

QUESTIONS FOR REVIEW

- 1 Explain how an increase in the price level affects the real value of money.
- 2 In one month, the CPI is reported as being 5.5 per cent and in the following month is reported as 5.0 per cent. Does this mean prices have fallen? Explain.
- 3 According to the quantity theory of money, what is the effect of an increase in the quantity of money?
- 4 Explain the difference between nominal and real variables, and give two examples of each. According to the principle of monetary neutrality, which variables are affected by changes in the quantity of money?
- 5 In what sense is inflation like a tax? How does thinking about inflation as a tax help explain hyperinflation?
- 6 What is meant by the velocity of money and what is its relevance to the Fisher equation?
- 7 According to the Fisher effect, how does an increase in the inflation rate affect the real interest rate and the nominal interest rate?
- 8 What are the costs of inflation? Which of these costs do you think are most important for your economy?
- 9 If inflation is less than expected, who benefits – debtors or creditors? Explain.
- 10 Why is deflation a potentially damaging phenomenon?

PROBLEMS AND APPLICATIONS

- 1 Suppose that this year's money supply is €500 billion, nominal GDP is €10 trillion and real GDP is €5 trillion.
 - a. What is the price level? What is the velocity of money?
 - b. Suppose that velocity is constant and the economy's output of goods and services rises by 5 per cent each year. What will happen to nominal GDP and the price level next year if the central bank keeps the money supply constant?
 - c. What money supply should the central bank set next year if it wants to keep the price level stable?
 - d. What money supply should the central bank set next year if it wants inflation of 10 per cent?
- 2 Suppose that changes in bank regulations expand the availability of credit cards, so that people need to hold less cash.
 - a. How does this event affect the demand for money?
 - b. If the central bank does not respond to this event, what will happen to the price level?
 - c. If the central bank wants to keep the price level stable, what should it do?
- 3 It is often suggested that central banks should try to achieve zero inflation. If we assume that velocity is constant, does this zero inflation goal require that the rate of money growth equal zero? If yes, explain why. If no, explain what the rate of money growth should equal.
- 4 The economist John Maynard Keynes wrote: 'Lenin is said to have declared that the best way to destroy the capitalist system was to debauch the currency. By a continuing process of inflation, governments can confiscate, secretly and unobserved, an important part of the wealth of their citizens.' Justify Lenin's assertion.
- 5 Suppose that a country's inflation rate increases sharply. What happens to the inflation tax on the holders of money? Why is wealth that is held in savings accounts not subject to a change in the inflation tax? Can you think of any way in which holders of savings accounts are hurt by the increase in the inflation rate?
- 6 Hyperinflations are extremely rare in countries whose central banks are independent of the rest of the government. Why might this be so?
- 7 Let's consider the effects of inflation in an economy composed only of two people: Michael, a bean farmer, and Dorothy, a rice farmer. Michael and Dorothy both always consume equal amounts of rice and beans. In year 2014, the price of beans was €1, and the price of rice was €3.
 - a. Suppose that in 2015 the price of beans was €2 and the price of rice was €6. What was inflation? Was Michael better off, worse off or unaffected by the changes in prices? What about Dorothy?
 - b. Now suppose that in 2015 the price of beans was €2 and the price of rice was €4. What was inflation? Was Michael better off, worse off or unaffected by the changes in prices? What about Dorothy?
 - c. Finally, suppose that in 2015 the price of beans was €2 and the price of rice was €1.50. What was inflation? Was Michael better off, worse off or unaffected by the changes in prices? What about Dorothy?
 - d. What matters more to Michael and Dorothy – the overall inflation rate or the relative price of rice and beans?
- 8 If the tax rate is 40 per cent, compute the before-tax real interest rate and the after-tax real interest rate in each of the following cases.
 - a. The nominal interest rate is 10 per cent and the inflation rate is 5 per cent.
 - b. The nominal interest rate is 6 per cent and the inflation rate is 2 per cent.
 - c. The nominal interest rate is 4 per cent and the inflation rate is 1 per cent.
- 9 Suppose that people expect inflation to equal 3 per cent, but in fact prices rise by 5 per cent. Describe how this unexpectedly high inflation rate would help or hurt the following:
 - a. the government
 - b. a homeowner with a fixed-rate mortgage
 - c. a union worker in the second year of a labour contract
 - d. a retired person who has invested their savings in government bonds.
- 10 Explain whether the following statements are true, false or uncertain.
 - a. 'Inflation hurts borrowers and helps lenders, because borrowers must pay a higher rate of interest.'
 - b. 'If prices change in a way that leaves the overall price level unchanged, then no one is made better or worse off.'
 - c. 'Inflation does not reduce the purchasing power of most workers.'

PART 13

THE MACROECONOMICS OF OPEN ECONOMIES

28

OPEN-ECONOMY MACROECONOMICS: BASIC CONCEPTS

When you next buy some fruit in the supermarket, the chances are that you will have a choice between a domestically produced fruit – perhaps apples – and fruit produced abroad, such as mangoes or bananas. When you take your next holiday, you may consider spending it in one of the cultural capitals of Europe or taking a trip to Disney World in Florida. When you start saving for your retirement, you may choose between a unit trust that buys mainly shares in domestic companies or one that buys shares of US or Japanese companies instead. In all of these cases, you are participating not just in the economy of your own country but in economies around the world.

There are clear benefits to being open to international trade: trade allows people to produce what they produce best and to consume the great variety of goods and services produced around the world. Indeed, one of the *Ten Principles of Economics* is that trade can make everyone better off. International trade can raise living standards in all countries by allowing each country to specialize in producing those goods and services in which it has a comparative advantage.

So far our development of macroeconomics has largely ignored the economy's interaction with other economies around the world. For most questions in macroeconomics, international issues are peripheral. For instance, when we discussed the natural rate of unemployment and the causes of inflation, the effects of international trade could safely be ignored. Indeed, to keep their analysis simple, macroeconomists often assume a closed economy – an economy that does not interact with other economies.

Yet some new macroeconomic issues arise in an open economy – an economy that interacts freely with other economies around the world. This chapter and the next one, therefore, provide an introduction to open-economy macroeconomics. We begin in this chapter by discussing the key macroeconomic variables that describe an open economy's interactions in world markets. You may have noticed mention of these variables – exports, imports, the trade balance and exchange rates – when reading the newspaper or watching the evening news. Our first job is to understand what these data mean. In the next chapter we develop a model to explain how these variables are determined and how they are affected by various government policies.

THE INTERNATIONAL FLOWS OF GOODS AND CAPITAL

An open economy interacts with other economies in two ways: it buys and sells goods and services in world product markets, and it buys and sells capital assets such as stocks and bonds in world financial markets. Here we discuss these two activities and the close relationship between them.

The Flow of Goods and Services: Exports, Imports and Net Exports

Exports are domestically produced goods and services that are sold abroad, and imports are foreign-produced goods and services that are sold domestically. When Lloyd's of London insures a building in Munich, it is paid an insurance premium for this service by the owner of the building. The sale of the insurance service provided by Lloyd's is an export for the United Kingdom and an import for Germany. When Volvo, the Swedish car manufacturer, makes a car and sells it to a Swiss resident, the sale is an import for Switzerland and an export for Sweden.

The net exports of any country are the value of its exports minus the value of its imports. The sale of insurance services abroad by Lloyd's raises UK net exports, and the Volvo sale reduces Swiss net exports. Because net exports tell us whether a country is, in total, a seller or a buyer in world markets for goods and services, net exports are also called the **trade balance**. If net exports are positive, exports are greater than imports, indicating that the country sells more goods and services abroad than it buys from other countries. In this case, the country is said to run a **trade surplus**. If net exports are negative, exports are less than imports, indicating that the country sells fewer goods and services abroad than it buys from other countries. In this case, the country is said to run a **trade deficit**. If net exports are zero, its exports and imports are exactly equal, and the country is said to have **balanced trade**.

trade balance the value of a nation's exports minus the value of its imports; also called net exports
trade surplus an excess of exports over imports
trade deficit an excess of imports over exports
balanced trade a situation in which exports equal imports

In the next chapter we develop a theory that explains an economy's trade balance, but even at this early stage it is easy to think of many factors that might influence a country's exports, imports and net exports. Those factors include the following:

- The tastes of consumers for domestic and foreign goods.
- The prices of goods at home and abroad.
- The exchange rates at which people can use domestic currency to buy foreign currencies.
- The incomes of consumers at home and abroad.
- The cost of transporting goods from country to country.
- The policies of the government towards international trade.

As these variables change over time, so does the amount of international trade.

The Flow of Financial Resources: Net Capital Outflow

So far we have been discussing how residents of an open economy participate in world markets for goods and services. In addition, residents of an open economy participate in world financial markets. A UK resident with £20,000 could use that money to buy a car from BMW, but he could instead use that money to buy stock in the German BMW corporation. The first transaction would represent a flow of goods, whereas the second would represent a flow of capital.

The term **net capital outflow** refers to the purchase of foreign assets by domestic residents minus the purchase of domestic assets by foreigners. (It is sometimes called *net foreign investment*.) When a

UK resident buys shares in BMW, the purchase raises UK net capital outflow. When a Japanese resident buys a bond issued by the UK government, the purchase reduces UK net capital outflow.

net capital outflow the purchase of foreign assets by domestic residents minus the purchase of domestic assets by foreigners

Recall that the flow of capital abroad takes two forms. If the French car manufacturer Renault opens up a factory in Romania, that is an example of *foreign direct investment*. Alternatively, if a French citizen buys shares in a Romanian company, that is an example of *foreign portfolio investment*. In the first case, the French owner is actively managing the investment, whereas in the second case the French owner has a more passive role. In both cases, French residents are buying assets located in another country, so both purchases increase French net capital outflow.

We develop a theory to explain net capital outflow in the next chapter. Here, let's consider briefly some of the more important variables that influence net capital outflow:

- The real interest rates being paid on foreign assets.
- The real interest rates being paid on domestic assets.
- The perceived economic and political risks of holding assets abroad.
- The government policies that affect foreign ownership of domestic assets.

For example, consider German investors deciding whether to buy Mexican government bonds or German government bonds. (Recall that a bond is, in effect, an IOU of the issuer.) To make this decision, German investors compare the real interest rates offered on the two bonds. The higher a bond's real interest rate, the more attractive it is. While making this comparison, however, German investors must also take into account the risk that one of these governments might *default* on its debt (that is, not pay interest or principal when it is due), as well as any restrictions that the Mexican government has imposed, or might impose in the future, on foreign investors in Mexico.

The Equality of Net Exports and Net Capital Outflow

We have seen that an open economy interacts with the rest of the world in two ways – in world markets for goods and services and in world financial markets. Net exports and net capital outflow each measure a type of imbalance in these markets. Net exports measure an imbalance between a country's exports and its imports. Net capital outflow measures an imbalance between the amount of foreign assets bought by domestic residents and the amount of domestic assets bought by foreigners.

An important but subtle fact of accounting states that, for an economy as a whole, these two imbalances must offset each other. That is, net capital outflow (*NCO*) always equals net exports (*NX*):

$$NCO = NX$$

This equation holds because every transaction that affects one side of this equation must also affect the other side by exactly the same amount. This equation is an *identity* – an equation that must hold because of the way the variables in the equation are defined and measured.

To see why this accounting identity is true, consider an example. Suppose that BP sells some aircraft fuel to a Japanese airline. In this sale, a UK company (BP) gives aircraft fuel to a Japanese company, and a Japanese company gives yen to a UK company. Notice that two things have occurred simultaneously. The United Kingdom has sold to a foreigner some of its output (the fuel), and this sale increases UK net exports. In addition, the United Kingdom has acquired some foreign assets (the yen), and this acquisition increases UK net capital outflow.

Although BP most probably will not hold on to the yen it has acquired in this sale, any subsequent transaction will preserve the equality of net exports and net capital outflow. For example, BP may exchange

its yen for pounds with a UK investment fund that wants the yen to buy shares in Sony Corporation, the Japanese maker of consumer electronics. In this case, BP's net export of aircraft fuel equals the investment fund's net capital outflow in Sony shares. Hence, NX and NCO rise by an equal amount.

Alternatively, BP may exchange its yen for pounds with another UK company that wants to buy computers from Toshiba, the Japanese computer maker. In this case, UK imports (of computers) exactly offset UK exports (of aircraft fuel). The sales by BP and Toshiba together affect neither UK net exports nor UK net capital outflow. That is, NX and NCO are the same as they were before these transactions took place.

The equality of net exports and net capital outflow follows from the fact that every international transaction is an exchange. When a seller country transfers a good or service to a buyer country, the buyer country gives up some asset to pay for this good or service. The value of that asset equals the value of the good or service sold. When we add everything up, the net value of goods and services sold by a country (NX) must equal the net value of assets acquired (NCO). The international flow of goods and services and the international flow of capital are two sides of the same coin.

Saving and Investment, and their Relationship to the International Flows

A nation's saving and investment are, as we have seen in earlier chapters, crucial to its long-run economic growth. Let's therefore consider how these variables are related to the international flows of goods and capital as measured by net exports and net capital outflow. We can do this most easily with the help of some simple mathematics.

As you may recall, the term net exports first appeared earlier in the book when we discussed the components of gross domestic product. The economy's gross domestic product (Y) is divided among four components: consumption (C), investment (I), government purchases (G) and net exports (NX). We write this as:

$$Y = C + I + G + NX$$

Total expenditure on the economy's output of goods and services is the sum of expenditure on consumption, investment, government purchases and net exports. Because each pound or euro of expenditure is placed into one of these four components, this equation is an accounting identity: it must be true because of the way the variables are defined and measured.

Recall that national saving is the income of the nation that is left after paying for current consumption and government purchases. National saving (S) equals $Y - C - G$. If we rearrange the above equation to reflect this fact, we obtain:

$$Y - C - G = I + NX$$

$$S = I + NX$$

Because net exports (NX) also equal net capital outflow (NCO), we can write this equation as:

$$S = I + NCO$$

| | | | | |
|----------|---|------------|---|-------------|
| S | = | I | + | NCO |
| Saving | | Domestic | | Net capital |
| | | investment | | outflow |

This equation shows that a nation's saving must equal its domestic investment plus its net capital outflow. In other words, when UK citizens save a pound of their income for the future, that pound can be used to finance accumulation of domestic capital or it can be used to finance the purchase of capital abroad.

This equation should look somewhat familiar. Earlier in the book, when we analysed the role of the financial system, we considered this identity for the special case of a closed economy. In a closed economy, net capital outflow is zero ($NCO = 0$), so saving equals investment ($S = I$). In contrast, an open economy has two uses for its saving: domestic investment and net capital outflow.

As before, we can view the financial system as standing between the two sides of this identity. For example, suppose the Singh family decides to save some of its income for retirement. This decision contributes to national saving, the left-hand side of our equation. If the Singh's deposit their saving in an investment fund, the fund may use some of the deposit to buy shares issued by BP, which uses the proceeds to build

an oil refinery in Aberdeen. In addition, the investment fund may use some of the Singh's deposit to buy shares issued by Toyota, which uses the proceeds to build a factory in Osaka. These transactions show up on the right-hand side of the equation. From the standpoint of UK accounting, the BP expenditure on a new oil refinery is domestic investment, and the purchase of Toyota stock by a UK resident is net capital outflow. Thus, all saving in the UK economy shows up as investment in the UK economy or as UK net capital outflow.

Summing Up

Table 28.1 summarizes many of the ideas presented so far in this chapter and will be important for later chapters. It describes the three possibilities for an open economy: a country with a trade deficit, a country with balanced trade and a country with a trade surplus.

TABLE 28.1 International Flows of Goods and Capital: Summary

This table shows the three possible outcomes for an open economy.

| Trade deficit | Balanced trade | Trade surplus |
|-------------------------|-------------------------|-------------------------|
| Exports < Imports | Exports = Imports | Exports > Imports |
| Net exports < 0 | Net exports = 0 | Net exports > 0 |
| $Y < C + I + G$ | $Y = C + I + G$ | $Y > C + I + G$ |
| Saving < Investment | Saving = Investment | Saving > Investment |
| Net capital outflow < 0 | Net capital outflow = 0 | Net capital outflow > 0 |

Consider first a country with a trade surplus. By definition, a trade surplus means that the value of exports exceeds the value of imports. Because net exports are exports minus imports, net exports (NX) are greater than zero. As a result, income ($Y = C + I + G + NX$) must be greater than domestic spending ($C + I + G$). But if Y is more than spending $C + I + G$, then saving $S = Y - C - G$ must be more than investment, I . Because the country is saving more than it is investing, it must be sending some of its saving abroad. That is, the net capital outflow must be greater than zero.

The converse logic applies to a country with a trade deficit. By definition, a trade deficit means that the value of exports is less than the value of imports. Because net exports are exports minus imports, net exports (NX) are negative. Thus, income ($Y = C + I + G + NX$) must be less than domestic spending ($C + I + G$). But if Y is less than $C + I + G$, then $S = Y - C - G$ must be less than I . That is, saving must be less than investment. The net capital outflow must be negative.

A country with balanced trade is between these cases. Exports equal imports, so net exports are zero. Income equals domestic spending, and saving equals investment. The net capital outflow equals zero.

SELF TEST Define net exports and net capital outflow. • Explain how they are related.

THE PRICES FOR INTERNATIONAL TRANSACTIONS: REAL AND NOMINAL EXCHANGE RATES

So far we have discussed measures of the flow of goods and services and the flow of capital across a nation's border. In addition to these quantity variables, macroeconomists also study variables that measure the prices at which these international transactions take place. Just as the price in any market serves the important role of coordinating buyers and sellers in that market, international prices help coordinate the decisions of consumers and producers as they interact in world markets. Here we discuss the two most important international prices – the nominal and real exchange rates.

Nominal Exchange Rates

The **nominal exchange rate** is the rate at which a person can trade the currency of one country for the currency of another. For example, if you go to a bank, you might see a posted exchange rate of 125 yen per euro. If you give the bank one euro, it will give you 125 Japanese yen; and if you give the bank 125 Japanese yen, it will give you one euro. (In fact, the bank will post slightly different prices for buying and selling yen. The difference gives the bank some profit for offering this service. For our purposes here, we can ignore these differences.)

nominal exchange rate the rate at which a person can trade the currency of one country for the currency of another

An exchange rate can always be expressed in two ways. If the exchange rate is 125 yen per euro, it is also $\frac{1}{125}$ (= 0.008) euro per yen. If a euro is worth £0.88, a pound is worth $\frac{1}{0.88}$ (= 1.136) euros. This can be a source of confusion, and there is no real hard and fast convention that people use. For example, it is customary to quote the US dollar–pound exchange rate as dollars per pound, e.g. \$1.50 if 1 pound exchanges for 1.50 dollars. On the other hand, the pound–euro exchange rate can be quoted either way, as pounds per euro or euros per pound. In this book we shall for the most part think of the exchange rate as being the quantity of foreign currency that exchanges for one unit of domestic currency, or the foreign price of a unit of domestic currency. For example, if we are thinking of the UK as the domestic economy and the USA as the foreign economy, then the exchange rate is \$1.50 per pound. If we are thinking of, say, Germany as the domestic economy, then we could express the exchange rate as dollars per euro, e.g. \$1.33 dollars per euro.

If the exchange rate changes so that a euro buys more of another currency, that change is called an **appreciation** of the euro. If the exchange rate changes so that a euro buys less of another currency, that change is called a **depreciation** of the euro. For example, when the exchange rate rises from 125 to 127 yen per euro, the euro is said to appreciate (you get a larger amount of yen in return for every euro). At the same time, because a Japanese yen now buys less of the European currency, the yen is said to depreciate (you have to give up more yen to buy one euro). When the exchange rate falls from 125 to 123 yen per euro, the euro is said to depreciate, and the yen is said to appreciate. (It is sometimes helpful to think how much of the domestic currency an individual has to give up to get the required amount of the foreign currency and vice versa.)

appreciation an increase in the value of a currency as measured by the amount of foreign currency it can buy
depreciation a decrease in the value of a currency as measured by the amount of foreign currency it can buy

At times you may have heard the media report that the pound or the euro is either ‘strong’ or ‘weak’. These descriptions usually refer to recent changes in the nominal exchange rate. When a currency appreciates, it is said to *strengthen* because it can then buy more foreign currency. Similarly, when a currency depreciates, it is said to *weaken*. If the individual gets more of the foreign currency in exchange for the same amount of the domestic currency, the domestic currency is stronger. If the individual has to give up more of the domestic currency to get the same amount of the foreign currency then the domestic currency is weaker.

For any currency, there are many nominal exchange rates. The euro can be used to buy US dollars, Japanese yen, British pounds, Mexican pesos and so on. When economists study changes in the exchange rate, they often use indices that average these many exchange rates. Just as the consumer prices index turns the many prices in the economy into a single measure of the price level, an exchange rate index turns these many exchange rates into a single measure of the international value of the currency. So when economists talk about the euro or the pound appreciating or depreciating, they often are referring to an exchange rate index that takes into account many individual exchange rates.

Real Exchange Rates

The **real exchange rate** is the rate at which a person can trade the goods and services of one country for the goods and services of another. For example, suppose that you go shopping and find that a kilo of Swiss cheese is twice as expensive as a kilo of English cheddar cheese. We would then say that the real exchange rate is a $\frac{1}{2}$ kilo of Swiss cheese per kilo of English cheese. Notice that, like the nominal exchange rate, we express the real exchange rate as units of the foreign item per unit of the domestic item. But in this instance the item is a good rather than a currency.

real exchange rate the rate at which a person can trade the goods and services of one country for the goods and services of another

Real and nominal exchange rates are closely related. To see how, consider an example. Suppose that a kilo of British wheat sells for £1 and a kilo of European wheat sells for €3. What is the real exchange rate between British and European wheat? To answer this question, we must first use the nominal exchange rate to convert the prices into a common currency. If the nominal exchange rate is €2 per pound, then a price for British wheat of £1 per kilo is equivalent to €2 per kilo. European wheat, however, sells for €3 a kilo, so British wheat is only $\frac{2}{3}$ as expensive as European wheat. The real exchange rate is $\frac{2}{3}$ of a kilo of European wheat per kilo of British wheat.

We can summarize this calculation for the real exchange rate with the following formula, where we are measuring the exchange rate as the amount of foreign currency needed to buy 1 unit of domestic currency:

$$\text{Real exchange rate} = \frac{(\text{Nominal exchange rate} \times \text{Domestic price})}{(\text{Foreign price})}$$

Using the numbers in our example, the formula applies as follows:

$$\begin{aligned} \text{Real exchange rate} &= \frac{(\text{€2 per pound}) \times (\text{€1 per kilo of UK wheat})}{\text{€3 per kilo of European wheat}} \\ &= \frac{\text{€2 per kilo of UK wheat}}{\text{€3 per kilo of European wheat}} \\ &= \frac{2}{3} \text{ kilo of European wheat per kilo of UK wheat} \end{aligned}$$

Thus, the real exchange rate depends on the nominal exchange rate and on the prices of goods in the two countries measured in the local currencies.

Why does the real exchange rate matter? As you might guess, the real exchange rate is a key determinant of how much a country exports and imports. For example, when a British bread company is deciding whether to buy British or European wheat to make into flour and use in making its bread, it will ask which wheat is cheaper. The real exchange rate gives the answer. As another example, imagine that you are deciding whether to take a holiday in the Dordogne, France, or in Cancun, Mexico. You might ask your travel agent the price of a hotel room in the Dordogne (measured in euros), the price of a hotel room in Cancun (measured in pesos) and the exchange rate between pesos and euros. If you decide where to go on holiday by comparing costs, you are basing your decision on the real exchange rate.

When studying an economy as a whole, macroeconomists focus on overall prices rather than the prices of individual items. That is, to measure the real exchange rate, they use price indices, such as the consumer prices index, which measure the price of a basket of goods and services. By using a price index for a UK or European basket (P), a price index for a foreign basket (P^*) and the nominal exchange rate between the UK pound or euro and foreign currencies (e = foreign currency per pound or euro), we can compute the overall real exchange rate between the United Kingdom or Europe and other countries as follows:

$$\text{Real exchange rate} = (e \times P)/P^*$$

This real exchange rate measures the price of a basket of goods and services available domestically relative to a basket of goods and services available abroad.

As we examine more fully in the next chapter, a country's real exchange rate is a key determinant of its net exports of goods and services. A depreciation (fall) in the real exchange rate of the euro means that EU goods have become cheaper relative to foreign goods. This change encourages consumers both at home and abroad to buy more EU goods and fewer goods from other countries. As a result, EU exports rise and EU imports fall, and both of these changes raise EU net exports. Conversely, an appreciation (rise) in the euro real exchange rate means that EU goods have become more expensive compared to foreign goods, so EU net exports fall. It is important to remember that whilst we are talking about the prices of exports and imports changing, the domestic price for these goods and services may not change. For example, a French wine producer may have wine for sale priced at €10 per bottle. If the exchange rate between the euro and the UK pound is £1 = €1.2 then a UK buyer of wine will have to give up £8.33 to buy a bottle of wine $\left(\frac{€10}{1.2}\right)$. If the UK exchange rate appreciates to £1 = €1.4 then the UK buyer now has to give up only £7.14 $\left(\frac{€10}{1.4}\right)$ to buy the bottle of wine. The euro price of the wine has not changed but to the UK buyer the price has fallen. Equally, if the pound exchange rate depreciated from £1 = €1.2 to £1 = €1.00 then the UK buyer would now have to give up £10 to buy the wine. Again, the euro price of the wine has not changed but the price to the UK buyer has risen because the exchange rate between the pound and the euro has changed.

SELF TEST Define nominal exchange rate and real exchange rate, and explain how they are related. • If the nominal exchange rate goes from 100 to 120 yen per euro, has the euro appreciated or depreciated?

A FIRST THEORY OF EXCHANGE RATE DETERMINATION: PURCHASING POWER PARITY

Exchange rates vary substantially over time. In 1970, one UK pound could buy 2.4 US dollars (i.e. the pound–dollar exchange rate was \$2.40), but in 1985 the pound was only worth about half this amount of dollars (the exchange rate was about \$1.25), in March 2008 one pound could buy over \$2 but by March 2013 the rate stood at around £1 = \$1.50. So over this 40-year period, the pound first almost halved in value from \$2.40 to \$1.25 and then increased by over 50 per cent from \$1.25 to over \$2. On the other hand, if we consider the value of the US dollar against the German and Italian currencies, we see that in 1970, a US dollar could be used to buy 3.65 German marks or 627 Italian lira, while in 1998, as both Germany and Italy were getting ready to adopt the euro as their common currency, a US dollar bought 1.76 German marks or 1,737 Italian lira. In other words, over this period the value of the dollar fell by more than half compared to the mark, while it more than doubled compared to the lira.

What explains these large changes? Economists have developed many models to explain how exchange rates are determined, each emphasizing just some of the many forces at work. Here we develop the simplest theory of exchange rates, called purchasing power parity (PPP). This theory states that a unit of any given currency should be able to buy the same quantity of goods in all countries. Many economists believe that **purchasing power parity** describes the forces that determine exchange rates in the long run. We now consider the logic on which this long-run theory of exchange rates is based, as well as the theory's implications and limitations.

purchasing power parity a theory of exchange rates whereby a unit of any given currency should be able to buy the same quantity of goods in all countries

The Basic Logic of Purchasing Power Parity

The theory of purchasing power parity is based on a principle called the *law of one price*. This law asserts that a good must sell for the same price in all locations. Otherwise, there would be opportunities for profit left unexploited. For example, suppose that coffee beans sold for less in Munich than in Frankfurt. A person could buy coffee in Munich for, say, €4 a kilo and then sell it in Frankfurt for €5 a kilo, making a profit of €1 per kilo from the difference in price. The process of taking advantage of differences in prices in different markets is called **arbitrage**. In our example, as the coffee buyer took advantage of the arbitrage opportunity, it would increase the demand for coffee in Munich and increase the supply in Frankfurt. The price of coffee would rise in Munich (in response to greater demand) and fall in Frankfurt (in response to greater supply). This process would continue until, eventually, the prices were the same in the two markets.

arbitrage a trade which seeks to exploit price differences in order to make profit

Now consider how the law of one price applies to the international marketplace. If a euro (or any other currency) could buy more coffee in Germany than in Japan, international traders could profit by buying coffee in Germany and selling it in Japan. This export of coffee from Germany to Japan would drive up the German price of coffee and drive down the Japanese price. Conversely, if a euro could buy more coffee in Japan than in Germany, traders could buy coffee in Japan and sell it in Germany. This import of coffee into Germany from Japan would drive down the German price of coffee and drive up the Japanese price. In the end, the law of one price tells us that a euro must buy the same amount of coffee in all countries.

This logic leads us to the theory of purchasing power parity. According to this theory, a currency must have the same purchasing power in all countries. That is, a euro must buy the same quantity of goods in Germany and Japan, and a Japanese yen must buy the same quantity of goods in Japan as in Germany. Indeed, the name of this theory describes it well. *Parity* means equality, and *purchasing power* refers to the value of money. *Purchasing power parity* states that a unit of all currencies must have the same real value in every country.

Implications of Purchasing Power Parity

What does the theory of purchasing power parity say about exchange rates? It tells us that the nominal exchange rate between the currencies of two countries depends on the price levels in those countries. If a euro buys the same quantity of goods in Germany (where prices are measured in euros) as in Japan (where prices are measured in yen), then the number of yen per euro must reflect the prices of goods in Germany and Japan. For example, if a kilo of coffee is priced at 500 yen in Japan and €5 in Germany, then the nominal exchange rate must be 100 yen per euro $\left(\frac{500 \text{ yen}}{\text{€}5 = 100 \text{ yen per euro}}\right)$. Otherwise, the purchasing power of the euro would not be the same in the two countries.

To see more fully how this works, it is helpful to use just a little mathematics. Think of Germany as the home or domestic economy. Suppose that P is the price of a basket of goods in Germany (measured in euros), P^* is the price of a basket of goods in Japan (measured in yen) and e is the nominal exchange rate (the number of yen needed to buy one euro). Now consider the quantity of goods a euro can buy at home (in Germany) and abroad. At home, the price level is P , so the purchasing power of €1 at home is $1/P$. Abroad, a euro can be exchanged into e units of foreign currency, which in turn have purchasing power e/P^* . For the purchasing power of a euro to be the same in the two countries, it must be the case that:

$$1/P = e/P^*$$

With rearrangement, this equation becomes:

$$1 = eP/P^*$$

Notice that the left-hand side of this equation is a constant, and the right-hand side is the real exchange rate. Thus, if the purchasing power of the euro is always the same at home and abroad, then the real exchange rate – the relative price of domestic and foreign goods – cannot change.

To see the implication of this analysis for the nominal exchange rate, we can rearrange the last equation to solve for the nominal exchange rate:

$$e = P^*/P$$

That is, the nominal exchange rate equals the ratio of the foreign price level (measured in units of the foreign currency) to the domestic price level (measured in units of the domestic currency). According to the theory of purchasing power parity, the nominal exchange rate between the currencies of two countries must reflect the different price levels in those countries.

A key implication of this theory is that nominal exchange rates change when price levels change. As we have seen earlier in the book, the price level in any country adjusts to bring the quantity of money supplied and the quantity of money demanded into balance. Because the nominal exchange rate depends on the price levels, it also depends on the money supply and money demand in each country. When a central bank in any country increases the money supply and causes the price level to rise, it also causes that country's currency to depreciate relative to other currencies in the world. In other words, when the central bank prints large quantities of money, that money loses value both in terms of the goods and services it can buy and in terms of the amount of other currencies it can buy.

Referring back to the beginning of this section, why did the UK pound lose value compared to the US dollar between 1970 and 1985? A good deal of the answer certainly relates to differences in inflation between the two countries. Between 1970 and 1985, the USA pursued, on average, a less inflationary monetary policy than the United Kingdom. Average price inflation in the UK over these 15 years was very high – about 10.5 per cent a year, while in the United States it was only about 6.5 per cent a year on average. This meant that between 1970 and 1985 the UK price level rose an average of 4 per cent a year faster than the US price level. As UK prices rose relative to US prices, the value of the pound fell relative to the dollar.

What about the movements in the exchange rate between the Italian lira and the dollar or the German mark and the dollar? Again, in large measure these movements reflect the relative stance of monetary policy in the countries concerned: Germany pursued a less inflationary monetary policy than the United States, and Italy pursued a more inflationary monetary policy. From 1970 to 1998, inflation in the United States was about 5.3 per cent per year on average. By contrast, average inflation was 3.5 per cent in Germany and 9.6 per cent in Italy. As US prices rose relative to German prices, the value of the dollar fell relative to the mark. Similarly, as US prices fell relative to Italian prices, the value of the dollar rose relative to the lira.

Germany and Italy are now part of a **common currency area**, and the common currency is the euro. This means that the two countries share a single monetary policy and that the inflation rates in the two countries will be closely linked. But the historical lessons of the lira and the mark will apply to the euro as well. Whether the UK pound or the US dollar buy more or fewer euros 20 years from now than they do today depends on whether the European Central Bank oversees more or less inflation in Europe than the Bank of England does in the United Kingdom or the Federal Reserve does in the United States.

common currency area a geographical area, possibly covering several countries, in which a common currency is used

FYI



Purchasing Power Standard (PPS)

If you look at statistics produced by Eurostat, the EU's statistical office, you are likely to see data expressed in *purchasing power standard (PPS)*. This is an extension of the PPP idea but applied to the EU. PPS is an artificial currency which expresses the purchasing power of the EU 28 against the euro. In theory, therefore, one PPS would buy the same amount of goods in each EU country and so $1\text{ PPS} = \text{€}1$. In reality there will be price differences between

European countries so different amounts of euro will be needed to buy the same goods and services in different countries. PPS takes account of these differences. The PPS is found by taking any national economic aggregate, such as GDP, for example, and dividing by its PPP. Thus PPPs can be seen as the exchange rate of the PPS against the euro and allows for more accurate comparisons between data from different EU countries. The PPS for

the EU 28 as a whole would be 100. If GDP per capita was expressed in PPS, then any figure above 100 would show GDP per capita in that country above the EU 28 average and any figure below 100 would indicate GDP per capita below the EU average. In 2011, Switzerland had GDP per capita at 157 and Norway was 186, both considerably higher than the EU 27 average at the time, whereas Poland at 64 and Romania at 49 were well below the EU 27 average.

Limitations of Purchasing Power Parity

Purchasing power parity provides a simple model of how exchange rates are determined. For understanding many economic phenomena, the theory works well. In particular, it can explain many long-term trends, such as the examples discussed earlier. It can also explain the major changes in exchange rates that occur during hyperinflations.

Yet the theory of purchasing power parity is not completely accurate. That is, exchange rates do not always move to ensure that a euro has the same real value in all countries all the time. There are two reasons why the theory of purchasing power parity does not always hold in practice.

The first reason is that many goods are not easily traded. Imagine, for instance, that haircuts are more expensive in Paris than in New York. International travellers might avoid getting their haircuts in Paris, and some haircutters might move from New York to Paris. Yet such arbitrage would probably be too limited to eliminate the differences in prices. Thus, the deviation from purchasing power parity might persist, and a euro (or dollar) would continue to buy less of a haircut in Paris than in New York.

The second reason that purchasing power parity does not always hold is that even tradable goods are not always perfect substitutes when they are produced in different countries. For example, some consumers prefer German cars, and others prefer Japanese cars. Moreover, consumer tastes can change over time. If German cars suddenly become more popular, the increase in demand will drive up the price of German cars compared to Japanese cars. But despite this difference in prices in the two markets, there might be no opportunity for profitable arbitrage because consumers do not view the two cars as equivalent.

Thus, both because some goods are not tradable and because some tradable goods are not perfect substitutes with their foreign counterparts, purchasing power parity is not a perfect theory of exchange rate determination. For these reasons, real exchange rates fluctuate over time. Nonetheless, the theory of purchasing power parity does provide a useful first step in understanding exchange rates. The basic logic is persuasive: as the real exchange rate drifts from the level predicted by purchasing power parity, people have greater incentive to move goods across national borders. Even if the forces of purchasing power parity do not completely fix the real exchange rate, they provide a reason to expect that changes in the real exchange rate are most often small or temporary. As a result, large and persistent movements in nominal exchange rates typically reflect changes in price levels at home and abroad.

CASE STUDY The Hamburger Standard

When economists apply the theory of purchasing power parity to explain exchange rates, they need data on the prices of a basket of goods available in different countries. One analysis of this sort is conducted by *The Economist*. This newspaper occasionally collects data on a basket of goods consisting of ‘two all-beef patties, special sauce, lettuce, cheese, pickles, onions, on a sesame seed bun.’ It’s called the ‘Big Mac’ and is sold by McDonald’s around the world.

Once we have the prices of Big Macs in two countries denominated in the local currencies, we can compute the exchange rate predicted by the theory of purchasing power parity. The predicted exchange rate is the one that makes the price of the Big Mac the same in the two countries. For example, assume that the price of a Big Mac in the United States is \$3.57 and in the UK, a Big Mac is £2.29. Purchasing power parity would predict an exchange rate of £1 = \$1.56.

$$\begin{aligned} \text{PPP}_{\text{£/\$}} &= \frac{\text{Price in dollars}}{\text{Price in pounds}} \\ \text{PPP}_{\text{£/\$}} &= \frac{3.57}{2.29} \\ \text{PPP}_{\text{£/\$}} &= 1.56 \text{ (rounded up)} \end{aligned}$$

If, however, the actual exchange rate quoted on the foreign exchanges was £1 = \$1.60 then the pound would be overvalued – you have to pay 4 cents more to acquire each pound than PPP implies. The extent of the overvaluation can be expressed as a percentage:

$$\begin{aligned} \% \text{ over or undervaluation} &= \frac{\text{Implied PPP rate} - \text{Market rate}}{\text{Market rate}} \times 100 \\ &= \left[\frac{(1.56 - 1.60)}{1.60} \right] \times 100 \\ &= \left(\frac{-0.04}{1.6} \right) \times 100 \\ &= -2.5\% \end{aligned}$$

In this example, the pound is overvalued against the dollar by 2.5 per cent and we might expect the pound to depreciate in the future.

If the market exchange rate was £1 = \$1.40, then the pound would be undervalued.

$$\begin{aligned} &= \left[\frac{(1.56 - 1.40)}{1.40} \right] \times 100 \\ &= \left(\frac{0.16}{1.4} \right) \times 100 \\ &= 11.43\% \end{aligned}$$

In this example, you have to pay 16 cents less than the PPP implies to buy a pound. It might be expected that the pound would appreciate in the future.

How well does purchasing power parity work when applied using Big Mac prices? Here are some examples from January 2013, when the price of a Big Mac was \$4.37 in the United States. The price of a Big Mac in China at this time was yuan 16.00. The implied PPP exchange rate therefore would be \$1 = yuan 3.66. The actual exchange rate was \$1 = yuan 6.22 which suggests that the yuan was undervalued by 41.1 per cent. Table 28.2 shows similar information for a range of other countries.

TABLE 28.2 Purchasing Parity Using the Big Mac Index

| Country | Price of a Big Mac | Actual exchange rate | Predicted exchange rate | US Dollar under or overvalued |
|----------------|--------------------|----------------------|-------------------------|-------------------------------|
| Chile | 2,050 pesos | 471.75 US\$/peso | 469.39 US\$/peso | Undervalued by 0.5% |
| United Kingdom | £2.69 | 0.63 US\$/pound | 0.62 US\$/pound | Undervalued by 2.7% |
| Sweden | 48.40 kronor | 6.35 US\$/kronor | 11.08 US\$/kronor | Overvalued by 74.5% |
| Russia | Roubles 72.88 | 30.05 US\$/Roubles | 16.69 US\$/Roubles | Undervalued by 44.5% |
| Japan | 320 yen | 91.06 US\$/yen | 73.27 US\$/yen | Undervalued by 19.5% |
| Mexico | 37 pesos | 12.74 US\$/peso | 8.47 US\$/pesos | Undervalued by 33.5% |
| Norway | 43.00 Krona | 5.48 US\$/krona | 9.85 US\$/krona | Overvalued by 79.6% |

You can see that the predicted and actual exchange rates are not exactly the same. After all, international arbitrage in Big Macs is not easy.

Purchasing power parity is not a precise theory of exchange rates (and Big Mac purchasing power parity even less so) and for some countries it does not always work well (it always tends to do badly, for example, for Scandinavian countries), but nevertheless it often provides a reasonable first approximation.



Wherever you go in the world the Big Mac is the same. This provides the standard for applying purchasing power parity to compute the Big Mac Index.

SELF TEST Over the past 20 years, Venezuela has had high inflation and Japan has had low inflation. What do you predict has happened to the number of Venezuelan bolivars a person can buy with a Japanese yen?

CONCLUSION

The purpose of this chapter has been to develop some basic concepts that macroeconomists use to study open economies. You should now understand why a nation's net exports must equal its net capital outflow, and why national saving must equal domestic investment plus net capital outflow. You should also understand the meaning of the nominal and real exchange rates, as well as the implications and limitations of purchasing power parity as a theory of how exchange rates are determined.

The macroeconomic variables defined here offer a starting point for analysing an open economy's interactions with the rest of the world. In the next chapter we develop a model that can explain what determines these variables. We can then discuss how various events and policies affect a country's trade balance and the rate at which nations make exchanges in world markets.

IN THE NEWS



Exchange Rates

One of the reasons for analysing macroeconomics from the perspective of an open economy is that decisions made within a country by policymakers can have repercussions across markets which affect different people in different ways.

Monetary Policy, Quantitative Easing and Exchange Rates

In the early part of 2013, the UK government faced a number of economic challenges, not least trying to reduce borrowing and the deficit, whilst at the same time facing the possibility of economic growth shrinking leading to a third period of recession in three years. Confidence in the UK economy at this time had an effect on the exchange rate. If investors feel that the UK economy is going to shrink there is an incentive to sell pounds on the foreign exchanges and buy a currency that gives a better return. What this does is cause the pound to depreciate.

The UK central bank, the Bank of England, meanwhile is tasked with keeping inflation under control at a target level of 2.0 per cent. In January 2013, the consumer prices index (CPI) was 2.7 per cent, the fourth month in a row that it had stood at this level. The Bank of England had noted that it expected inflation to accelerate in 2013. In minutes of the meeting of the Bank of England's Monetary Policy Committee (MPC), published in February 2013, it became clear that the MPC had discussed cutting interest rates further although in the end the MPC voted to keep the repo rate at 0.5 per cent. There were also discussions about the possibility of extending quantitative easing. In the event, the MPC voted not to expand QE but, given

this was not the first time the issue had been discussed, the markets made an assumption that it was highly likely that it was just a matter of time before QE was expanded further. If the Bank of England extends its asset purchasing facility then the demand for bonds increases and as prices rise, yields fall (remember there is an inverse relationship between the price and the yield of a bond). Falling interest rates do not provide much of an incentive for foreign investors to put their money into the UK.

Then, on 22 February 2013, it was announced that one of the ratings agencies that values credit worthiness had cut the UK government's rating from the highest level of AAA to AA1. The agency,

Moody's, said that the economic climate was still highly uncertain and that the government's debt reduction measures still had a long way to go.

All this news compounded the lack of confidence in the UK economy and led to a further depreciation of the pound against key currencies, the US dollar and the euro. In the six months to February 2013, the pound fell by 10 per cent against the euro and was heading towards parity ($\text{£}1 = \text{€}1$). The pound had also slipped from over $\text{£}1 = \$1.60$ to around $\text{£}1 = \$1.52$ over this period. Analysts at the time

expected the value of the pound to depreciate further.

A declining value of the pound might reflect a lack of confidence in the UK economy by the financial markets but it could be good news for exporters. As sterling weakens against major trading nations in the EU and the United States, it means that exporters become more competitive and this could mean a boost in exports. This could help the UK economy to recover but exactly how far will depend on a variety of factors, not least the price elasticity of demand for exports.

On the other hand, firms importing raw materials and semi-finished products will see costs rising. Some might pass on those costs to consumers in the form of higher prices but in a difficult economic climate that might not be easy. For those importing raw materials and semi-finished products and then exporting the resulting finished goods, the effect will depend on the balance between the rise in import costs and the benefits from lower export prices.

Questions

- 1 It has been suggested that quantitative easing is a back-door method of manipulating currencies to help engineer an export-led recovery. Explain.
- 2 Use a supply and demand diagram to show the effect on the exchange rate between the UK pound and the euro of a decision by investors to pull out of UK investments.
- 3 Why would a cut in the UK government's credit rating lead to a depreciation of the pound?
- 4 Explain, using appropriate examples, how a depreciation of a currency affects exporters and importers and the relevance of the price elasticity of demand for imports and exports.
- 5 To what extent is a depreciation of the currency likely to help a country to turn economic stagnation into economic growth?

The Bank of England – sometimes referred to as the Grand Old Lady of Threadneedle Street – has to balance managing the need to maintain price stability and the external effects of its policies.



SUMMARY

- Net exports are the value of domestic goods and services sold abroad minus the value of foreign goods and services sold domestically. Net capital outflow is the acquisition of foreign assets by domestic residents minus the acquisition of domestic assets by foreigners. Because every international transaction involves an exchange of an asset for a good or service, an economy's net capital outflow always equals its net exports.
- An economy's saving can be used either to finance investment at home or to buy assets abroad. Thus, national saving equals domestic investment plus net capital outflow.
- The nominal exchange rate is the relative price of the currency of two countries, and the real exchange rate is the relative price of the goods and services of two countries. When the nominal exchange rate changes so that each unit of domestic currency buys more foreign currency, the domestic currency is said to *appreciate* or *strengthen*. When the nominal exchange rate changes so that each unit of domestic currency buys less foreign currency, the domestic currency is said to *depreciate* or *weaken*.
- According to the theory of purchasing power parity, a unit of currency should be able to buy the same quantity of goods in all countries. This theory implies that the nominal exchange rate between the currencies of two countries should reflect the price levels in those countries. As a result, countries with relatively high inflation should have depreciating currencies, and countries with relatively low inflation should have appreciating currencies.

QUESTIONS FOR REVIEW

- 1 Why do economists sometimes conduct analysis under assumptions of a closed economy?
- 2 What do you understand by the term an 'open economy'?
- 3 If a British citizen visits the French Alps to go for a skiing holiday, is this an import for France or an export? Explain.
- 4 What is a trade deficit and why is it important?
- 5 Define net exports and net capital outflow. Explain how and why they are related.
- 6 Explain the relationship between saving, investment and net capital outflow.
- 7 If a Japanese car is priced at 500,000 yen, a similar German car is priced at €10,000, and a euro can buy 100 yen, what are the nominal and real exchange rates?
- 8 Describe the economic logic behind the theory of purchasing power parity.
- 9 If the European Central Bank started printing large quantities of euros, what would happen to the number of Japanese yen a euro could buy?
- 10 Explain the effect of an appreciation of the euro on German exporters and German importers.

PROBLEMS AND APPLICATIONS

- 1 How would the following transactions affect UK exports, imports and net exports?
 - a. A British art lecturer spends the summer touring museums in Italy.
 - b. Students in Paris flock to see the latest Royal Shakespeare Company perform *King Lear* on tour.
 - c. The British art lecturer buys a new Volvo.
 - d. A student in Munich buys a Manchester United official team shirt (in Munich).
 - e. A British citizen goes to Calais for the day to stock up on cheap wine.
- 2 International trade in each of the following products has increased over time. Suggest some reasons why this might be so:
 - a. wheat
 - b. banking services
 - c. computer software
 - d. automobiles.
- 3 Describe the difference between foreign direct investment and foreign portfolio investment. Who is more likely to engage in foreign direct investment – a corporation or an individual investor? Who is more likely to engage in foreign portfolio investment?
- 4 How would the following transactions affect UK net capital outflow? Also, state whether each involves direct investment or portfolio investment.
 - a. A British mobile telephone company establishes an office in the Czech Republic.
 - b. A US company's pension fund buys shares in BP.
 - c. Toyota expands its factory in Derby, England.
 - d. A London-based investment trust sells its Volkswagen shares to a French investor.

- 5 Holding national saving constant, does an increase in net capital outflow increase, decrease or have no effect on a country's accumulation of domestic capital?
- 6 The business section of most major newspapers contains a table showing exchange rates amongst many countries, as does *The Economist*. Find such a table (or find *The Economist* table on its website) and use it to answer the following questions.
- Does this table show nominal or real exchange rates? Explain.
 - What are the exchange rates between the euro and the UK pound, and between the UK pound and the US dollar? Calculate the exchange rate between the euro and the dollar.
 - If UK inflation exceeds European inflation over the next year, would you expect the UK pound to appreciate or depreciate relative to the euro?
- 7 Would each of the following groups be happy or unhappy if the euro appreciated? Explain.
- US pension funds holding French government bonds.
 - German manufacturing industries.
 - Australian tourists planning a trip to Europe.
 - A British firm trying to purchase property overseas.
- 8 What is happening to the Swiss real exchange rate in each of the following situations? Explain.
- The Swiss nominal exchange rate is unchanged, but prices rise faster in Switzerland than abroad.
 - The Swiss nominal exchange rate is unchanged, but prices rise faster abroad than in Switzerland.
 - The Swiss nominal exchange rate declines, and prices are unchanged in Switzerland and abroad.
 - The Swiss nominal exchange rate declines, and prices rise faster abroad than in Switzerland.
- 9 List three goods for which the law of one price is likely to hold, and three goods for which it is not. Justify your choices.
- 10 A can of lemonade is priced at €0.75 in Europe and 12 pesos in Mexico. What would the peso–euro exchange rate be if purchasing power parity holds? If a monetary expansion caused all prices in Mexico to double, so that lemonade rose to 24 pesos, what would happen to the peso–euro exchange rate?

29

A MACROECONOMIC THEORY OF THE OPEN ECONOMY

Over the past two decades, the United Kingdom and United States have persistently imported more goods and services than they have exported. That is, UK and US net exports have been negative. Japan and Germany, on the other hand, have persistently exported more than they have imported (although since 2008 Japan has experienced negative net exports).

If the government of a country wants to reduce trade deficits, what should it do? Should it try to limit imports, perhaps by imposing a quota on the import of cars from Japan or Germany? Or should it try to influence the country's trade deficit in some other way?

To understand what factors determine a country's trade balance and how government policies can affect it, we need a macroeconomic theory of the open economy. The preceding chapter introduced some of the key macroeconomic variables that describe an economy's relationship with other economies – including net exports, net capital outflow, and the real and nominal exchange rates. This chapter develops a model that identifies the forces that determine these variables and shows how these variables are related to one another.

To develop this macroeconomic model of an open economy, we build on our previous analysis in two important ways. First, the model takes the economy's GDP as given. We assume that the economy's output of goods and services, as measured by real GDP, is determined by the supplies of the factors of production and by the available production technology that turns these inputs into output. Second, the model takes the economy's price level as given. We assume that the price level adjusts to bring the supply and demand for money into balance. In other words, this chapter takes as a starting point the lessons learned in Chapters 22 and 27 about the determination of the economy's output and price level.

The goal of the model in this chapter is to highlight the forces that determine the economy's trade balance and exchange rate. In one sense, the model is simple: it applies the tools of supply and demand to an open economy. Yet the model is also more complicated than others we have seen because it involves looking simultaneously at two related markets – the market for loanable funds and the market for foreign currency exchange. After we develop this model of the open economy, we use it to examine how various events and policies affect the economy's trade balance and exchange rate. We shall then be able to determine the government policies that are most likely to reverse trade deficits.

SUPPLY AND DEMAND FOR LOANABLE FUNDS AND FOR FOREIGN CURRENCY EXCHANGE

To understand the forces at work in an open economy, we focus on supply and demand in two markets. The first is the market for loanable funds, which coordinates the economy's saving, investment and the flow of loanable funds abroad (called the net capital outflow). The second is the market for foreign currency exchange, which coordinates people who want to exchange the domestic currency for the currency of other countries. In this section we discuss supply and demand in each of these markets. In the next section we put these markets together to explain the overall equilibrium for an open economy.

The Market for Loanable Funds

When we first analysed the role of the financial system in Chapter 23, we made the simplifying assumption that the financial system consists of only one market, called the *market for loanable funds*. All savers go to this market to deposit their saving and all borrowers go to this market to get their loans. In this market, there is one interest rate, which is both the return to saving and the cost of borrowing.

To understand the market for loanable funds in an open economy, the place to start is the identity discussed in the preceding chapter:

$$\begin{array}{rcl} S & = & I \quad + \quad NCO \\ \text{Saving} & = & \text{Domestic investment} \quad + \quad \text{Net capital outflow} \end{array}$$

Whenever a nation saves some of its income, it can use that money to finance the purchase of domestic capital or to finance the purchase of an asset abroad. The two sides of this identity represent the two sides of the market for loanable funds. The supply of loanable funds comes from national saving (S). The demand for loanable funds comes from domestic investment (I) and net capital outflow (NCO). Note that the purchase of a capital asset adds to the demand for loanable funds, regardless of whether that asset is located at home or abroad. Because net capital outflow can be either positive or negative, it can either add to or subtract from the demand for loanable funds that arises from domestic investment.

As we learned in Chapter 23, the quantity of loanable funds supplied and the quantity of loanable funds demanded depend on the real interest rate. A higher real interest rate encourages people to save and, therefore, raises the quantity of loanable funds supplied. A higher interest rate also makes borrowing to finance capital projects more costly; thus, it discourages investment and reduces the quantity of loanable funds demanded.

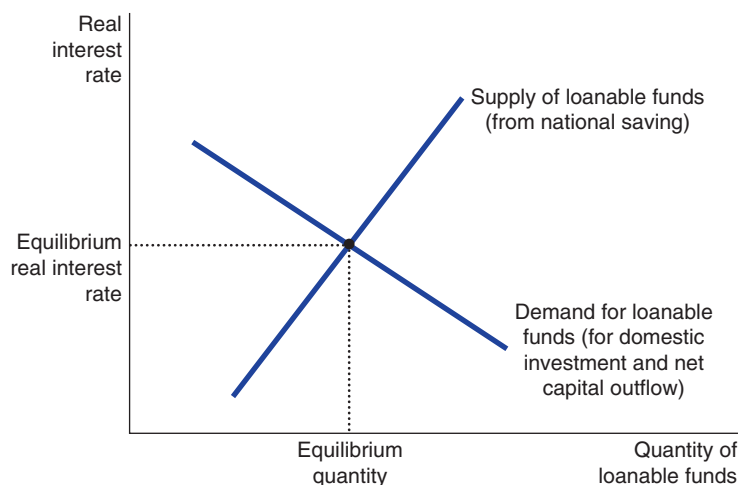
In addition to influencing national saving and domestic investment, the real interest rate in a country affects that country's net capital outflow. To see why, consider two investment funds – one in the United Kingdom and one in Germany – deciding whether to buy a UK government bond or a German government bond. The investment funds would make this decision in part by comparing the real interest rates in the United Kingdom and Germany. When the UK real interest rate rises, the UK bond becomes more attractive to both investment funds. Thus, an increase in the UK real interest rate discourages Brits from buying foreign assets and encourages people living in other countries to buy UK assets. For both reasons, a high UK real interest rate reduces UK net capital outflow.

We represent the market for loanable funds on the familiar supply-and-demand diagram in Figure 29.1. As in our earlier analysis of the financial system, the supply curve slopes upward because a higher interest rate increases the quantity of loanable funds supplied, and the demand curve slopes downward because a higher interest rate decreases the quantity of loanable funds demanded.

FIGURE 29.1

The Market for Loanable Funds

The interest rate in an open economy, as in a closed economy, is determined by the supply and demand for loanable funds. National saving is the source of the supply of loanable funds. Domestic investment and net capital outflow are the sources of the demand for loanable funds. At the equilibrium interest rate, the amount that people want to save exactly balances the amount that people want to borrow for the purpose of buying domestic capital and foreign assets.



Unlike the situation in our previous discussion, however, the demand side of the market now represents the behaviour of both domestic investment and net capital outflow. That is, in an open economy, the demand for loanable funds comes not only from those who want loanable funds to buy domestic capital goods but also from those who want loanable funds to buy foreign assets.

The interest rate adjusts to bring the supply and demand for loanable funds into balance. If the interest rate were below the equilibrium level, the quantity of loanable funds supplied would be less than the quantity demanded. The resulting shortage of loanable funds would push the interest rate upward. Conversely, if the interest rate were above the equilibrium level, the quantity of loanable funds supplied would exceed the quantity demanded. The surplus of loanable funds would drive the interest rate downward. At the equilibrium interest rate, the supply of loanable funds exactly balances the demand. That is, at the equilibrium interest rate, the amount that people want to save exactly balances the desired quantities of domestic investment and net capital outflow.

The Market for Foreign Currency Exchange

The second market in our model of the open economy is the market for foreign currency exchange. Let's think of the UK as the domestic economy. Participants in this market trade UK pounds in exchange for foreign currencies. To understand the market for foreign currency exchange, we begin with another identity from the last chapter:

$$\begin{array}{lcl} NCO & = & NX \\ \text{Net capital outflow} & = & \text{Net exports} \end{array}$$

This identity states that the imbalance between the purchase and sale of capital assets abroad (NCO) equals the imbalance between exports and imports of goods and services (NX). For example, when the UK economy is running a trade surplus ($NX > 0$), foreigners are buying more UK goods and services than UK residents are buying foreign goods and services. What are Brits doing with the foreign currency they are getting from this net sale of goods and services abroad? They must be buying foreign assets, so UK capital is flowing abroad ($NCO > 0$). Conversely, if the UK is running a trade deficit ($NX < 0$), Brits are spending more on foreign goods and services than they are earning from selling abroad. Some of this spending must be financed by selling UK assets abroad, so foreign capital is flowing into the UK ($NCO < 0$).

Our model of the open economy treats the two sides of this identity as representing the two sides of the market for foreign currency exchange. Net capital outflow represents the quantity of pounds supplied for the purpose of buying foreign assets. For example, when a UK investment fund wants to buy a Japanese government bond, it needs to change pounds into yen, so it supplies pounds in the market for foreign currency exchange. Net exports represent the quantity of pounds demanded for the purpose of buying UK net exports of goods and services. For example, when a Japanese airline wants to buy aircraft fuel produced by BP, it needs to change its yen into pounds, so it demands pounds in the market for foreign currency exchange.

What price balances the supply and demand in the market for foreign currency exchange? The answer is the real exchange rate. As we saw in the preceding chapter, the real exchange rate is the relative price of domestic and foreign goods and, therefore, is a key determinant of net exports. When the UK real exchange rate appreciates, UK goods become more expensive relative to foreign goods, making UK goods less attractive to consumers abroad (exports would rise) and foreign goods more attractive to domestic consumers (imports would rise). For both reasons, net UK exports fall. Hence, an appreciation of the real exchange rate reduces the quantity of pounds demanded in the market for foreign currency exchange.

Figure 29.2 shows supply and demand in the market for foreign currency exchange. The demand curve slopes downward for the reason we just discussed: a higher real exchange rate makes UK goods more expensive and reduces the quantity of pounds demanded to buy those goods. The supply curve is vertical because the quantity of pounds supplied for net capital outflow does not depend on the real exchange rate. (As discussed earlier, net capital outflow depends on the real interest rate. When discussing the market for foreign currency exchange, we take the real interest rate and net capital outflow as given.)

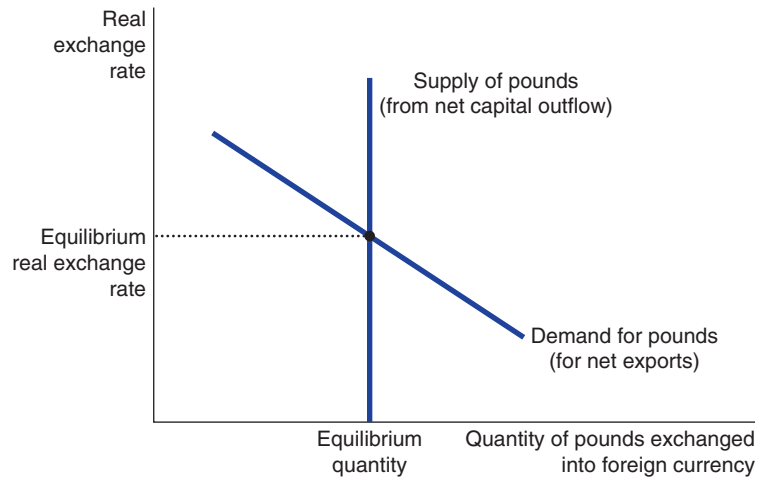
The real exchange rate adjusts to balance the supply and demand for pounds just as the price of any good adjusts to balance supply and demand for that good. If the real exchange rate were below the equilibrium level, the quantity of pounds supplied would be less than the quantity demanded. The resulting shortage of pounds would push the value of the pound upwards. Conversely, if the real exchange rate

were above the equilibrium level, the quantity of pounds supplied would exceed the quantity demanded. The surplus of pounds would drive the value of the pound downward. At the equilibrium real exchange rate, the demand for pounds by non-UK residents arising from the UK net exports of goods and services exactly balances the supply of pounds from UK residents arising from UK net capital outflow.

FIGURE 29.2

The Market for Foreign Currency Exchange

The real exchange rate is determined by the supply and demand for foreign currency exchange. The supply of pounds to be exchanged into foreign currency comes from net capital outflow. Because net capital outflow does not depend on the real exchange rate, the supply curve is vertical. The demand for pounds comes from net exports. Because a lower real exchange rate stimulates net exports (and thus increases the quantity of pounds demanded to pay for these net exports), the demand curve is downwards sloping. At the equilibrium real exchange rate, the number of pounds people supply to buy foreign assets exactly balances the number of pounds people demand to buy net exports.



At this point, it is worth noting that the division of transactions between ‘supply’ and ‘demand’ in this model is somewhat artificial. In our model, net exports are the source of the demand for pounds, and net capital outflow is the source of the supply. Thus, when a UK resident imports a car made in Japan, our model treats that transaction as a decrease in the quantity of pounds demanded (because net exports fall) rather than an increase in the quantity of pounds supplied. Similarly, when a Japanese citizen buys a UK government bond, our model treats that transaction as a decrease in the quantity of pounds supplied (because net capital outflow falls) rather than an increase in the quantity of pounds demanded. This use of language may seem somewhat unnatural at first, but it will prove useful when analysing the effects of various policies.

SELF TEST Describe the sources of supply and demand in the market for loanable funds and the market for foreign currency exchange.

EQUILIBRIUM IN THE OPEN ECONOMY

So far we have discussed supply and demand in two markets – the market for loanable funds and the market for foreign currency exchange. Let’s now consider how these markets are related to each other.

Net Capital Outflow: The Link Between the Two Markets

We begin by recapping what we’ve learned so far in this chapter. We have been discussing how the economy coordinates four important macroeconomic variables: national saving (S), domestic investment (I), net capital outflow (NCO) and net exports (NX). Keep in mind the following identities:

$$S = I + NCO$$

and

$$NCO = NX$$

FYI



Purchasing Power Parity as a Special Case

An alert reader of this book might ask: why are we developing a theory of the exchange rate here? Didn't we already do that in the preceding chapter?

As you may recall, the preceding chapter developed a theory of the exchange rate called purchasing power parity. This theory asserts that a euro (or any other currency) must buy the same quantity of goods and services in every country. As a result, the real exchange rate is fixed, and all changes in the nominal exchange rate between two currencies reflect changes in the price levels in the two countries.

The model of the exchange rate developed here is related to

the theory of purchasing power parity. According to the theory of purchasing power parity, international trade responds quickly to international price differences. If goods were cheaper in one country than in another, they would be exported from the first country and imported into the second until the price difference disappeared. In other words, the theory of purchasing power parity assumes that net exports are highly responsive to small changes in the real exchange rate. If net exports were in fact so responsive, the demand curve in Figure 29.2 would be horizontal.

Thus, the theory of purchasing power parity can be viewed as a special case of the model

considered here. In that special case, the demand curve for foreign currency exchange, rather than being downwards sloping, is horizontal at the level of the real exchange rate that ensures parity of purchasing power at home and abroad. That special case is a good place to start when studying exchange rates, but it is far from the end of the story.

This chapter, therefore, concentrates on the more realistic case in which the demand curve for foreign currency exchange is downwards sloping. This allows for the possibility that the real exchange rate changes over time, as in fact it sometimes does in the real world.

In the market for loanable funds, supply comes from national saving, demand comes from domestic investment and net capital outflow, and the real interest rate balances supply and demand. In the market for foreign currency exchange, supply comes from net capital outflow, demand comes from net exports, and the real exchange rate balances supply and demand.

Net capital outflow is the variable that links these two markets. In the market for loanable funds, net capital outflow is part of demand. A person who wants to buy an asset abroad must finance this purchase by obtaining resources in the market for loanable funds. In the market for foreign currency exchange, net capital outflow is the source of supply. A person who wants to buy an asset in another country must supply pounds in order to exchange them for the currency of that country.

The key determinant of net capital outflow, as we have discussed, is the real interest rate. When the UK interest rate is high, owning UK assets is more attractive, and UK net capital outflow is low. Figure 29.3 shows this negative relationship between the interest rate and net capital outflow. This net capital outflow curve is the link between the market for loanable funds and the market for foreign currency exchange.

Simultaneous Equilibrium in Two Markets

We can now put all the pieces of our model together in Figure 29.4. This figure shows how the market for loanable funds and the market for foreign currency exchange jointly determine the important macroeconomic variables of an open economy.

Panel (a) of the figure shows the market for loanable funds (taken from Figure 29.1). As before, national saving is the source of the supply of loanable funds. Domestic investment and net capital outflow are the source of the demand for loanable funds. The equilibrium real interest rate (r_1) brings the quantity of loanable funds supplied and the quantity of loanable funds demanded into balance.

FIGURE 29.3

How Net Capital Outflow Depends on the Interest Rate

Because a higher domestic real interest rate makes domestic assets more attractive, it reduces net capital outflow. Note the position of zero on the horizontal axis: net capital outflow can be positive or negative. A negative value of net capital outflow means that the economy is experiencing a net inflow of capital.

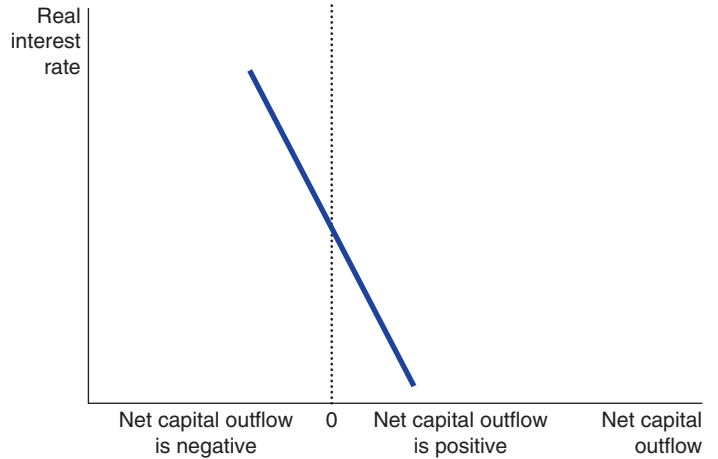
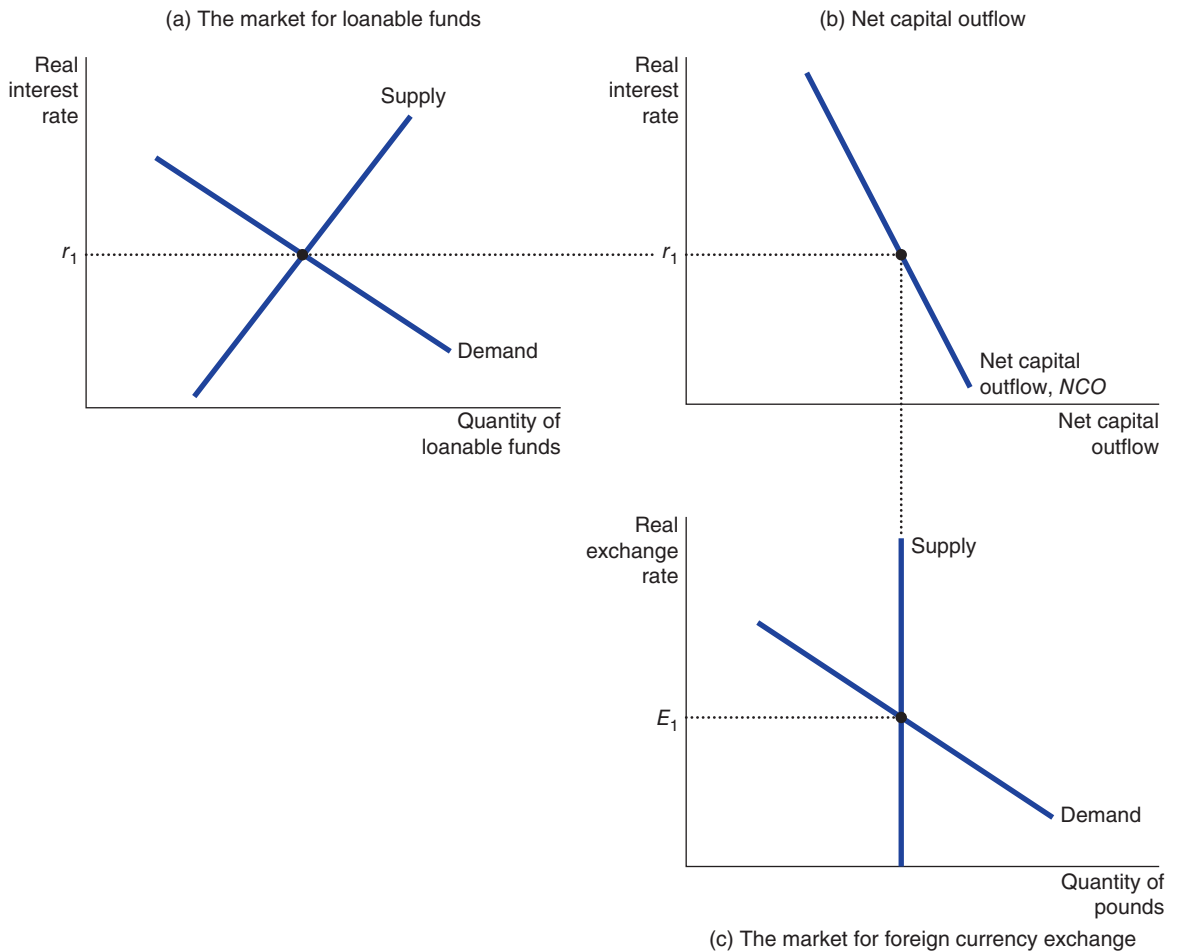


FIGURE 29.4

The Real Equilibrium in an Open Economy

In panel (a), the supply and demand for loanable funds determine the real interest rate. In panel (b), the interest rate determines net capital outflow, which provides the supply of pounds in the market for foreign currency exchange. In panel (c), the supply and demand for pounds in the market for foreign currency exchange determine the real exchange rate.



Panel (b) of the figure shows net capital outflow (taken from Figure 29.3). It shows how the interest rate from panel (a) determines net capital outflow. A higher interest rate at home makes domestic assets more attractive, and this in turn reduces net capital outflow. Therefore, the net capital outflow curve in panel (b) slopes downward.

Panel (c) of the figure shows the market for foreign currency exchange (taken from Figure 29.2). Because foreign assets must be purchased with foreign currency, the quantity of net capital outflow from panel (b) determines the supply of pounds to be exchanged into foreign currencies. The real exchange rate does not affect net capital outflow, so the supply curve is vertical. The demand for pounds comes from net exports. Because a depreciation of the real exchange rate increases net exports, the demand curve for foreign currency exchange slopes downward. The equilibrium real exchange rate (E_1) brings into balance the quantity of pounds supplied and the quantity of pounds demanded in the market for foreign currency exchange.

The two markets shown in Figure 29.4 determine two relative prices – the real interest rate and the real exchange rate. The real interest rate determined in panel (a) is the price of goods and services in the present relative to goods and services in the future. The real exchange rate determined in panel (c) is the price of domestic goods and services relative to foreign goods and services. These two relative prices adjust simultaneously to balance supply and demand in these two markets. As they do so, they determine national saving, domestic investment, net capital outflow and net exports. In a moment, we will use this model to see how all these variables change when some policy or event causes one of these curves to shift.

SELF TEST In the model of the open economy just developed, two markets determine two relative prices.
 • What are the markets? • What are the two relative prices?

HOW POLICIES AND EVENTS AFFECT AN OPEN ECONOMY

Having developed a model to explain how key macroeconomic variables are determined in an open economy, we can now use the model to analyse how changes in policy and other events alter the economy's equilibrium. As we proceed, keep in mind that our model is just supply and demand in two markets – the market for loanable funds and the market for foreign currency exchange and as such can be analysed using the three-step approach outlined in Chapter 3.

Government Budget Deficits

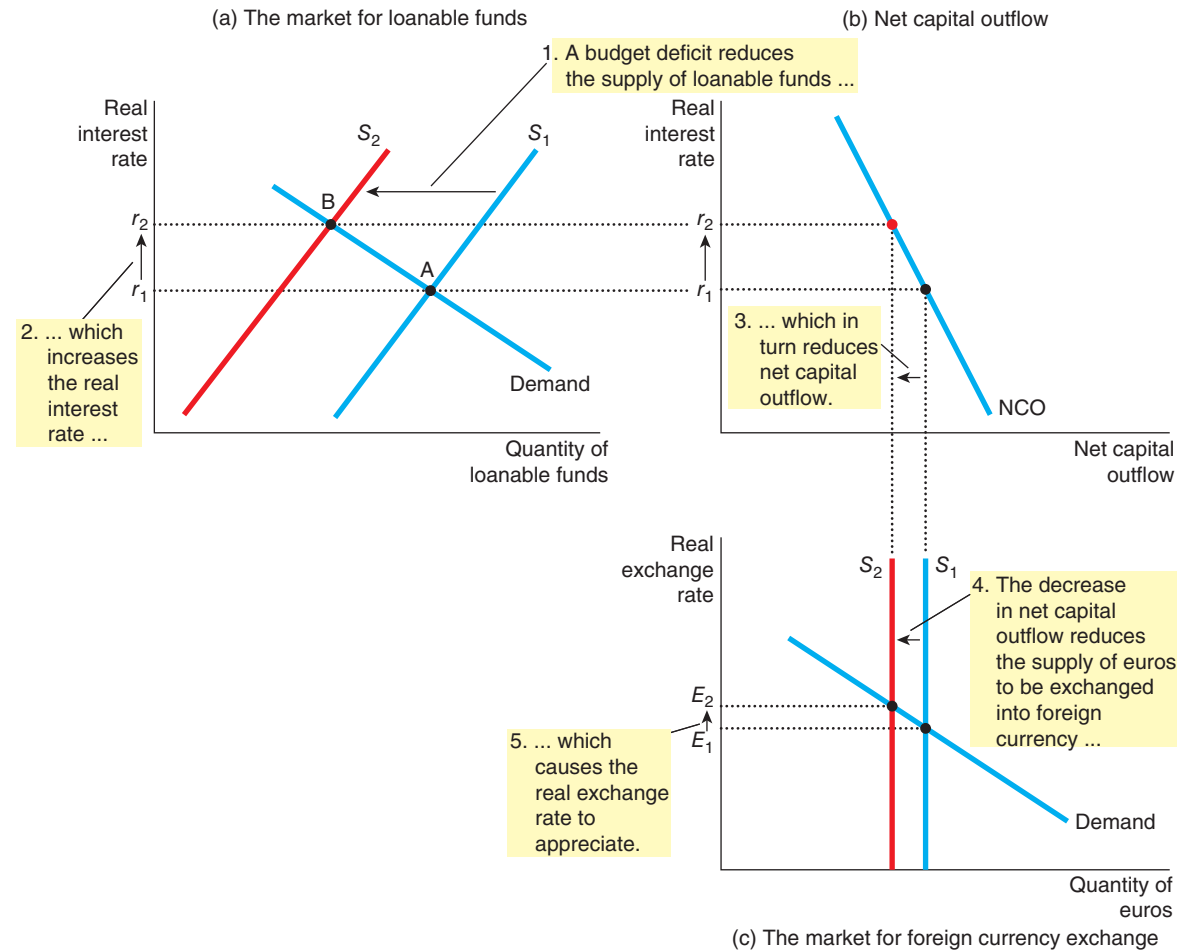
When we first discussed the supply and demand for loanable funds earlier in the book, we examined the effects of government budget deficits, which occur when government spending exceeds government revenue. Because a government budget deficit represents *negative* public saving, it reduces national saving (the sum of public and private saving). Thus, a government budget deficit reduces the supply of loanable funds, drives up the interest rate and crowds out investment.

Now let's consider the effects of a budget deficit in an open economy. First, which curve in our model shifts? As in a closed economy, the initial impact of the budget deficit is on national saving and, therefore, on the supply curve for loanable funds. Second, which way does this supply curve shift? Again as in a closed economy, a budget deficit represents *negative* public saving, so it reduces national saving and shifts the supply curve for loanable funds to the left. This is shown as the shift from S_1 to S_2 in panel (a) of Figure 29.5.

Our third and final step is to compare the old and new equilibria. Panel (a) shows the impact of a French budget deficit on the French market for loanable funds. With fewer funds available for borrowers in French financial markets, the interest rate rises from r_1 to r_2 to balance supply and demand.

FIGURE 29.5**The Effects of a Government Budget Deficit**

If the French government runs a budget deficit, it reduces the supply of loanable funds from S_1 to S_2 in panel (a). The interest rate rises from r_1 to r_2 to balance the supply and demand for loanable funds. In panel (b), the higher interest rate reduces net capital outflow. Reduced net capital outflow, in turn, reduces the supply of euros in the market for foreign currency exchange from S_1 to S_2 in panel (c). This fall in the supply of euros causes the real exchange rate to appreciate from E_1 to E_2 . The appreciation of the exchange rate pushes the trade balance towards deficit.



Faced with a higher interest rate, borrowers in the market for loanable funds choose to borrow less. This change is represented in the figure as the movement from point A to point B along the demand curve for loanable funds. In particular, households and firms reduce their purchases of capital goods. As in a closed economy, budget deficits crowd out domestic investment.

In an open economy, however, the reduced supply of loanable funds has additional effects. Panel (b) shows that the increase in the interest rate from r_1 to r_2 reduces net capital outflow. (This fall in net capital outflow is also part of the decrease in the quantity of loanable funds demanded in the movement from point A to point B in panel (a).) Because saving kept at home now earns higher rates of return, investing abroad is less attractive, and domestic residents buy fewer foreign assets. Higher interest rates also attract foreign investors, who want to earn the higher returns on French assets. Thus, when budget deficits raise interest rates, both domestic and foreign behaviour cause French net capital outflow to fall.

Panel (c) shows how budget deficits affect the market for foreign currency exchange. Because net capital outflow is reduced, people need less foreign currency to buy foreign assets, and this induces a leftward shift in the supply curve for euros from S_1 to S_2 . The reduced supply of euros causes the real exchange rate to appreciate from E_1 to E_2 . That is, the euro becomes more valuable compared to foreign currencies. This appreciation, in turn, makes French goods more expensive compared to foreign goods. Because people both at home and abroad switch their purchases away from the more expensive French goods, exports from France fall, and imports into France rise. For both reasons, French net exports fall. Hence, in an open economy, government budget deficits raise real interest rates, crowd out domestic investment, cause the currency to appreciate and push the trade balance toward deficit.

The budget deficit and trade deficit are so closely related in both theory and practice that, especially when they are both large, they are often referred to by the nickname of the *twin deficits*. We should not, however, view these twins as identical, for many factors beyond fiscal policy can influence the trade deficit.

Trade Policy

A **trade policy** is a government policy that directly influences the quantity of goods and services that a country imports or exports. Trade policy takes various forms. One common trade policy is a *tariff*, a tax on imported goods. Another is an *import quota*, a limit on the quantity of a good that can be produced abroad and sold domestically. Trade policies are common throughout the world, although sometimes they are disguised. For example, before 2000 there was an understanding between Japan and the European Union that Japan would voluntarily limit its sales of cars into the UK, France, Italy, Portugal and Spain to a maximum of 1.1 million (excluding cars produced at factories owned by Japanese companies but located within the European Union). These so-called ‘voluntary export restrictions’ are not really voluntary and, in essence, are a form of import quota.

trade policy a government policy that directly influences the quantity of goods and services that a country imports or exports

Let’s consider the macroeconomic impact of trade policy. Suppose that the European car industry, concerned about competition from Japanese car makers, convinces the European Union (EU) to impose a quota on the number of cars that can be imported from Japan into the EU. In making their case, lobbyists for the car industry assert that the trade restriction would improve the overall EU trade balance. Are they right? Our model, as illustrated in Figure 29.6, offers an answer.

The first step in analysing the trade policy is to determine which curve shifts. The initial impact of the import restriction is, not surprisingly, on imports. Because net exports equal exports minus imports, the policy also affects net exports. And because net exports are the source of demand for euros in the market for foreign currency exchange, the policy affects the demand curve in this market.

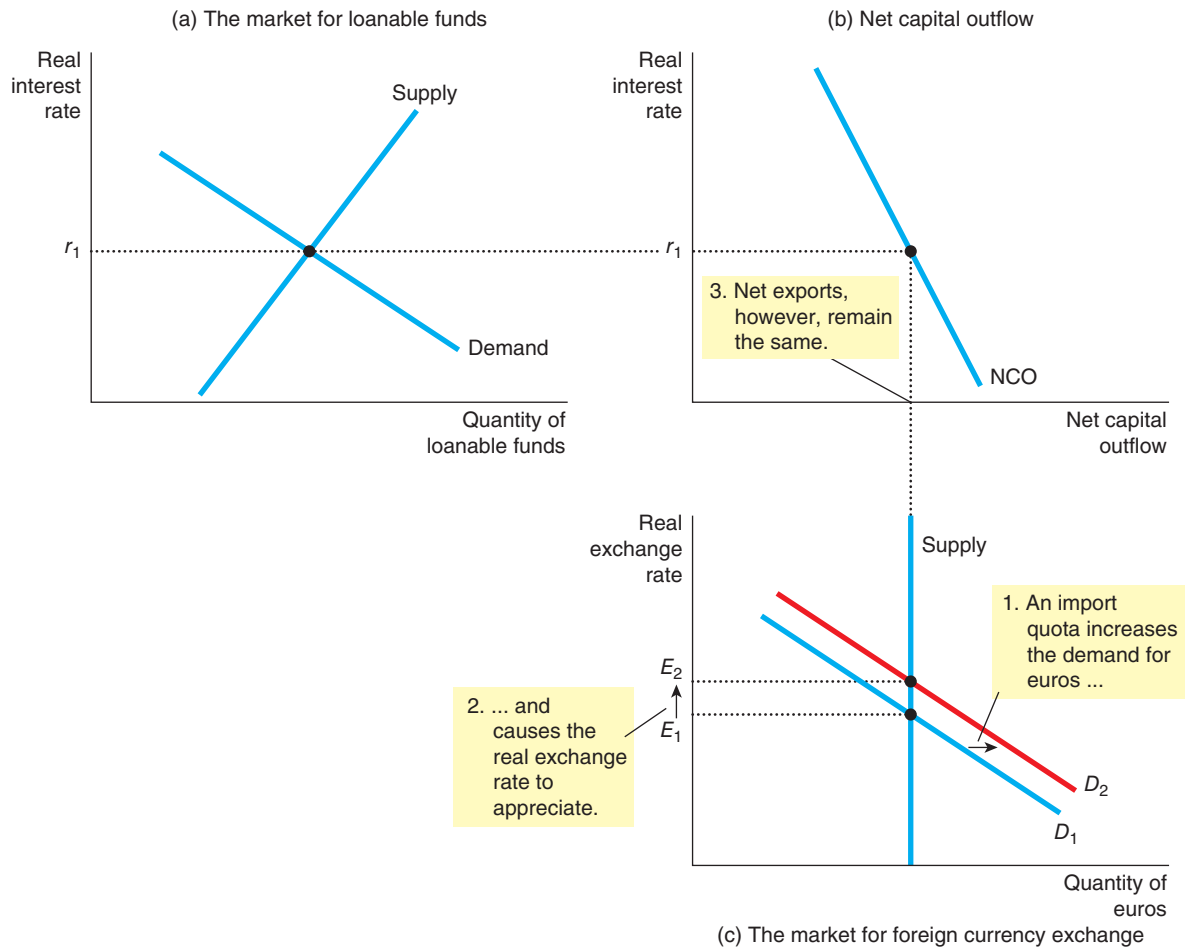
The second step is to determine which way this demand curve shifts. Because the quota restricts the number of Japanese cars sold in the EU, it reduces imports at any given real exchange rate. Net exports, which equal exports minus imports, will therefore *rise* for any given real exchange rate. Because non-Europeans need euros to buy EU net exports, there is an increased demand for euros in the market for foreign currency exchange. This increase in the demand for euros is shown in panel (c) of Figure 29.6 as the shift from D_1 to D_2 .

The third step is to compare the old and new equilibria. As we can see in panel (c), the increase in the demand for euros causes the real exchange rate to appreciate from E_1 to E_2 . Because nothing has happened in the market for loanable funds in panel (a), there is no change in the real interest rate. Because there is no change in the real interest rate, there is also no change in net capital outflow, shown in panel (b). And because there is no change in net capital outflow, there can be no change in net exports, even though the import quota has reduced imports.

FIGURE 29.6

The Effects of an Import Quota

When the EU imposes a quota on the import of Japanese cars, nothing happens in the market for loanable funds in panel (a) or to net capital outflow in panel (b). The only effect is a rise in net exports (exports minus imports) for any given real exchange rate. As a result, the demand for euros in the market for foreign currency exchange rises, as shown by the shift from D_1 to D_2 in panel (c). This increase in the demand for euros causes the value of the euro to appreciate from E_1 to E_2 . This appreciation of the euro tends to reduce net exports, offsetting the direct effect of the import quota on the trade balance.



The reason why net exports can stay the same while imports fall is explained by the change in the real exchange rate: when the euro appreciates in value in the market for foreign currency exchange, European goods become more expensive relative to non-European goods. This appreciation encourages imports and discourages exports – and both of these changes work to offset the direct increase in net exports due to the import quota. In the end, an import quota reduces both imports and exports, but net exports (exports minus imports) are unchanged.

We have thus come to a surprising implication: trade policies do not affect the trade balance. That is, policies that directly influence exports or imports do not alter net exports. This conclusion seems less surprising if one recalls the accounting identity:

$$NX = NCO = S - I$$

Net exports equal net capital outflow, which equals national saving minus domestic investment. Trade policies do not alter the trade balance because they do not alter national saving or domestic investment.

For given levels of national saving and domestic investment, the real exchange rate adjusts to keep the trade balance the same, regardless of the trade policies the government puts in place.

Although trade policies do not affect a country's overall trade balance, these policies do affect specific firms, industries and countries. When the EU imposes an import quota on Japanese cars, European car makers have less competition from abroad and will sell more cars. At the same time, because the euro has appreciated in value, Airbus, the European aircraft maker, will find it harder to compete with Boeing, the US aircraft maker. European exports of aircraft will fall, and European imports of aircraft will rise. In this case, the import quota on Japanese cars will increase net exports of cars and decrease net exports of aeroplanes. In addition, it will increase net exports from the EU to Japan and decrease net exports from the EU to the United States. The overall trade balance of the EU economy, however, stays the same.

The effects of trade policies are, therefore, more microeconomic than macroeconomic. Although advocates of trade policies sometimes claim (incorrectly) that these policies can alter a country's trade balance, they are usually more motivated by concerns about particular firms or industries. One should not be surprised, for instance, to hear an executive from BMW advocating import quotas for Japanese cars. Economists almost always oppose such trade policies. Free trade allows economies to specialize in doing what they do best, making residents of all countries better off. Trade restrictions interfere with these gains from trade and, thus, reduce overall economic well-being.

Capital Flight

Large and sudden movement of funds out of a country is called **capital flight**. To see the implications of capital flight for an economy, we again follow our three steps for analysing a change in equilibrium.

capital flight a large and sudden reduction in the demand for assets located in a country

Consider first which curves in our model capital flight affects. We will use France as an example here because in early 2013 it was reported that the country had experienced a considerable outflow of funds amounting to €53 billion in the two months to February 2013. If investors around the world decide to sell some of their assets in France and use the proceeds to buy assets in other countries, this increases French net capital outflow and, therefore, affects both markets in our model. Most obviously, it affects the net capital outflow curve, and this in turn influences the supply of euros in the market for foreign currency exchange. In addition, because the demand for loanable funds comes from both domestic investment and net capital outflow, capital flight affects the demand curve in the market for loanable funds.

Now consider which way these curves shift. When net capital outflow increases, there is greater demand for loanable funds to finance these purchases of capital assets abroad. Thus, as panel (a) of Figure 29.7 shows, the demand curve for loanable funds shifts to the right from D_1 to D_2 . In addition, because net capital outflow is higher for any interest rate, the net capital outflow curve also shifts to the right from NCO_1 to NCO_2 , as in panel (b).

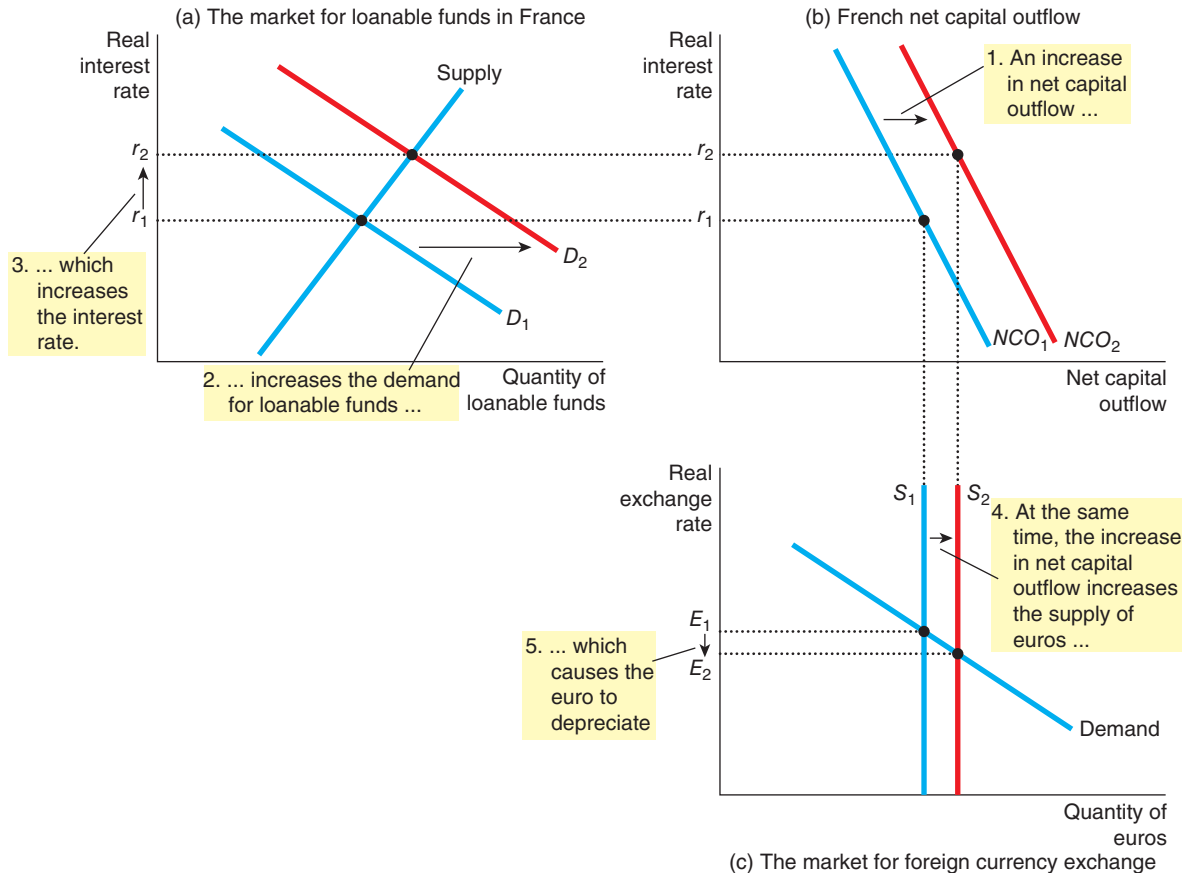
To see the effects of capital flight on the economy, we compare the old and new equilibria. Panel (a) of Figure 29.7 shows that the increased demand for loanable funds causes the interest rate in France to rise from r_1 to r_2 . Panel (b) shows that French net capital outflow increases. (Although the rise in the interest rate does make French assets more attractive, this only partly offsets the impact of capital flight on net capital outflow.) Panel (c) shows that the increase in net capital outflow raises the supply of euros in the market for foreign currency exchange from S_1 to S_2 . That is, as people try to get out of French assets, there is a large supply of euros to be converted into other currencies such as dollars and pounds. This increase in supply causes the euro to depreciate from E_1 to E_2 . Thus, capital flight from France increases French interest rates and decreases the value of the euro in the market for foreign currency exchange.

These price changes that result from capital flight influence some key macroeconomic quantities. The depreciation of the currency makes exports cheaper and imports more expensive, pushing the trade balance towards surplus. At the same time, the increase in the interest rate reduces domestic investment, which slows capital accumulation and economic growth.

FIGURE 29.7

The Effects of Capital Flight

If people decide that France is a risky place to keep their savings, they will move their capital to safer havens, resulting in an increase in French net capital outflow. Consequently, the demand for loanable funds in France rises from D_1 to D_2 , as shown in panel (a), and this drives up French real interest rate from r_1 to r_2 . Because net capital outflow is higher for any interest rate, that curve also shifts to the right from NCO_1 to NCO_2 in panel (b). At the same time, in the market for foreign currency exchange, the supply of euros rises from S_1 to S_2 , as shown in panel (c). This increase in the supply of euros causes the euro to depreciate from E_1 to E_2 , so the euro becomes less valuable compared to other currencies.



Although capital flight has its largest impact on the country from which capital is fleeing, it also affects other countries. When capital flows out of France into the United States, for instance, it has the opposite effect on the US economy as it has on the French economy. In particular, the rise in French net capital outflow coincides with a fall in US net capital outflow. As the euro depreciates in value and French interest rates rise, the dollar appreciates in value and US interest rates fall. The size of this impact on the US economy is dependent on the relative size of the economy of the United States compared to that of France, and in turn the extent to which the change in the supply of euros from France affects the euro exchange rate as a whole, given that France is just one of 18 countries which make up the euro area.

SELF TEST Suppose that Europeans decided to spend a smaller fraction of their incomes. • What would be the effect on saving, investment, interest rates, the real exchange rate and the trade balance in the European Union?

CONCLUSION

Historically, international trade has always played a very important role in most European economies. In the past two centuries or so, international finance has also become increasingly important. The typical modern European country consumes a high proportion of goods produced abroad and exports a significant amount of its output to other European countries and to countries outside Europe. In addition, through investment funds and other financial institutions, Europeans borrow and lend in world financial markets, as indeed do the citizens of all advanced industrialized economies.

It is clear, therefore, that a proper understanding of macroeconomics requires a study of the workings of the open economy. This chapter has provided a basic model for thinking about the macroeconomics of open economies.

IN THE NEWS

Capital Flight

Capital flight can be caused by a lack of confidence in an economy or by political instability but it can also manifest itself in funds leaving a country illegally and possibly linked to criminal activity.

Billions Leaving Russia Illegally

The economy of Russia has changed considerably since the early 1990s as the move to a more market-based economy has taken hold. However, the country does suffer from illegal flights of capital both in and out of the country. Estimates put the illegal capital flight since 1994 at some \$782.5 billion (€593.2 billion) and in early 2013 the governor of Russia's central bank, Sergei Ignatiev, said that some \$49 billion (€37.1 billion) had left the country illegally in 2012. To put this in context, the sum represents around 2.5 per cent of Russia's GDP.

The Russian government claimed that the figure was exaggerated but a study by a US research organization, Global Financial Integrity, in 2013 suggested that on average, some \$64 billion (€48.5 billion) had entered or left Russia every year since 2004. Much of this money, it claims, is

income generated from criminal activities which include weapons smuggling, prostitution, people trafficking, drugs and corruption.

The central bank governor noted that the figure he had announced was arrived at by looking at 'payments made by Russian organizations to non-residents, the stated aims of which are clearly false'. In addition, Mr Ignatiev noted that the country also faced problems from so-called 'one-day firms' which are commercial organizations set up to act as channels for illegal funds but which are closed down immediately the funds have been channelled and therefore avoid paying any taxes.

Finding ways to combat this illegal flight is one of the major challenges facing the Russian government. If flows of capital out of the country are as large as were being suggested then the opportunity cost to the country in terms of what the funds could have been used for – investment in infrastructure, health

and education – is extremely high. Some inside Russia have noted that if the country fails to get to grips

Money from criminal gangs involved with prostitution and drugs, among other things, may be part of the funds entering and leaving Russia every year – capital flight has a cost beyond the purely economic.



with the problem and reduces the incidences of corruption which are alleged to be at high levels within the government, the development of the economy would be held back.

Questions

- 1 Assuming the amount of money leaving Russia is greater than that coming in, what happens to the demand for loanable funds and the interest rate in Russia?
- 2 What effect would you expect capital flight of the size outlined in the article to have on the exchange rate between the Russian rouble and other currencies?
- 3 What effect is the capital flight likely to have on Russia's trading partners given the size of the sums quoted in the article?
- 4 Does the fact that the flows are illegal make any difference to your analysis in the previous questions?
- 5 Why is clamping down on illegal capital flight and corruption important to Russia's continued economic development?

SUMMARY

- To analyse the macroeconomics of open economies, two markets are central – the market for loanable funds and the market for foreign currency exchange. In the market for loanable funds, the interest rate adjusts to balance the supply of loanable funds (from national saving) and the demand for loanable funds (from domestic investment and net capital outflow). In the market for foreign currency exchange, the real exchange rate adjusts to balance the supply of domestic currency (for net capital outflow) and the demand for domestic currency (for net exports). Because net capital outflow is part of the demand for loanable funds and provides the supply of domestic currency for foreign currency exchange, it is the variable that connects these two markets.
- A policy that reduces national saving, such as a government budget deficit, reduces the supply of loanable funds and drives up the interest rate. The higher interest rate reduces net capital outflow, which reduces the supply of domestic currency in the market for foreign currency exchange. The domestic currency appreciates and net exports fall.
- Although restrictive trade policies, such as tariffs or quotas on imports, are sometimes advocated as a way to alter the trade balance, they do not necessarily have that effect. A trade restriction increases net exports for a given exchange rate and, therefore, increases the demand for the domestic currency in the market for foreign currency exchange. As a result, the domestic currency appreciates in value, making domestic goods more expensive relative to foreign goods. This appreciation offsets the initial impact of the trade restriction on net exports.
- When investors change their attitudes about holding assets of a country, the ramifications for the country's economy can be profound. In particular, political instability or a lack of confidence in an economy can lead to capital flight, which tends to increase interest rates and causes the currency to depreciate.

QUESTIONS FOR REVIEW

- 1 Why do we describe the formula $S = I + NCO$ as 'an identity'?
- 2 How is the interest rate in an open economy determined?
- 3 Describe supply and demand in the market for loanable funds and the market for foreign currency exchange. How are these markets linked?
- 4 What effect does a government budget deficit have on the interest rate? Draw a diagram to illustrate your answer.
- 5 What effect would you expect a government budget surplus to have on the exchange rate? Explain.
- 6 Why are budget deficits and trade deficits sometimes called the twin deficits?
- 7 What is 'trade policy'?
- 8 Suppose that a steel workers' union encourages people to buy only European steel. What would this policy do to the European overall trade balance and the real exchange rate? What is the impact on the European steel industry? What is the impact on the European textile industry?
- 9 What effect would you expect a tariff to have on the real exchange rate? Explain using appropriate diagrams.
- 10 What is capital flight? When a country experiences capital flight, what is the effect on its interest rate and exchange rate?

PROBLEMS AND APPLICATIONS

- 1 Germany generally runs a trade surplus. Do you think this is most related to high foreign demand for German goods, low German demand for foreign goods, a high German saving rate relative to German investment, or structural barriers against imports into Germany? Explain your answer.
- 2 If the UK prime minister announces that the government are solidly on a course of deficit reduction, which should make the pound more attractive to investors, would such a deficit reduction in fact raise the value of the pound? Explain.
- 3 Suppose that the government passes an investment tax credit, which subsidizes domestic investment. How does this policy affect national saving, domestic investment, net capital outflow, the interest rate, the exchange rate and the trade balance?
- 4 Assume that there is a rise in the trade deficit of a country due largely to the rise in a government budget deficit. Assume also that some commentators in the popular press claim that the increased trade deficit resulted from a decline in the quality of the country's products relative to foreign products.
 - a. Assume that the country's products did decline in relative quality during this period. How might this affect net exports *at any given exchange rate*?
 - b. Use a three-panel diagram to show the effect of this shift in net exports on the country's real exchange rate and trade balance.
 - c. Does a decline in the quality of the country's products have any effect on standards of living for its residents? (Hint: when a country's residents sell goods to non-country residents, what do they receive in return?)
- 5 Explain in words why European *export* industries would benefit from a reduction in restrictions on *imports* into the European Union.
- 6 Suppose the French suddenly develop a strong taste for British wine. Answer the following questions in words and using a diagram.
 - a. What happens to the demand for pounds in the market for foreign currency exchange?
 - b. What happens to the value of pounds in the market for foreign currency exchange?
 - c. What happens to the quantity of UK net exports?
- 7 Suppose your country is running a trade deficit and you hear your trade minister on the radio, saying: 'The trade deficit must be reduced, but import quotas only annoy our trading partners. If we subsidize our exports instead, we can reduce the deficit by increasing our competitiveness.' Using a three-panel diagram, show the effect of an export subsidy on net exports and the real exchange rate. Do you agree with the trade minister?
- 8 Suppose that real interest rates increase across Europe. Explain how this development will affect UK net capital outflow. Then explain how it will affect UK net exports by using a formula from the chapter and by using a diagram. What will happen to the UK real interest rate and real exchange rate?
- 9 Suppose that Germans decide to increase their saving.
 - a. If the elasticity of German net capital outflow with respect to the real interest rate is very high, will this increase in private saving have a large or small effect on German domestic investment?
 - b. If the elasticity of German exports with respect to the real exchange rate is very low, will this increase in private saving have a large or small effect on the German real exchange rate?
- 10 Assume that saving in China has been used to finance investment into the EU. That is, the Chinese have been buying European capital assets.
 - a. If the Chinese decided they no longer wanted to buy European assets, what would happen in the European market for loanable funds? In particular, what would happen to European interest rates, European saving and European investment?
 - b. What would happen in the market for foreign currency exchange? In particular, what would happen to the value of the euro and the European trade balance?

PART 14

SHORT-RUN ECONOMIC FLUCTUATIONS

30 BUSINESS CYCLES

The majority of you reading this chapter will have lived through a period of economic change. Many of you will have been born in the mid-1990s. Between this time and the middle part of the first decade of the 2000s, many countries in Europe experienced a period of prosperity, relatively strong economic growth, relatively low and stable inflation, unemployment rates which were not too high and interest rates which were at relatively low levels.

From about 2007 onwards, things changed. The financial crisis led to serious economic problems in many countries and a global recession. A **recession** is characterized by a period of falling incomes and rising unemployment and technically occurs after two successive quarters of negative growth. If such a contraction in growth continues and is more severe it might be described as a **depression**.

recession a period of declining real incomes and rising unemployment. The technical definition gives recession occurring after two successive quarters of negative economic growth

depression a severe recession

Whilst many countries came out of recession around 2009–2010, economic growth in many European countries has been weak. Interest rates in many developed countries have been at virtually zero rates for a number of years, unemployment has risen to alarming levels in some countries but inflation has been stuck at levels above the target rate for many countries. At the time of writing, the economic prospects do not look good but by the time you read this chapter things may be improving in your country. By 2016 it could be that many economies in Europe return once again to a period of stronger growth and an increase in prosperity.

Your instructors will have experienced the ups and downs of economic activity over a longer period but they too will have experienced periods of considerable economic problems and periods of strong growth. The latter part of the 1970s and into the early part of the 1980s was characterized by severe economic problems in the UK but by the mid-1980s, things improved only to slump again in the early 1990s.

It is clear that the level of economic activity, the amount of buying and selling (the number of transactions) changes over time. Economic activity is an important factor in determining economic growth and in

turn the level of employment, unemployment, inflation and other macroeconomic variables. Economists are pretty much agreed on the causes of long-run changes in macroeconomic equilibrium but there is more disagreement about what causes short-run variations in economic activity. The short-run swings in economic activity vary in size and time-span. In the UK, the recession of 1920–1924, for example, saw GDP fall by about 9 per cent from its peak over the first 18 months and then take a further 27 months to get back to its pre-recession peak, and in the next 12 months rise to around 4 per cent above its pre-recession peak. In 1930–1934, the length of time for the economy to recover to its pre-recession peak was around 48 months but the economy ‘only’ fell to around 7 per cent below its pre-recession peak. In the 1990–1993 recession, the economy took around 32 months to recover but thereafter rose by over 8 per cent above its pre-recession peak and only fell to around 3 per cent of its pre-recession peak. The post-financial crisis of 2008 until the time of writing has seen the economy falling to around 6 per cent below its pre-recession peak and has been below the peak consistently for over 60 months. You may be reading this now and the UK economy has still not managed to reach its pre-recession peak. The **business cycle** refers to the study of the fluctuations in economic growth around the trend growth.

business cycle the fluctuations in economic growth around trend growth

TREND GROWTH RATES

Most of the data we will be looking at in this chapter is classed as **time series data**, observations on a variable over a time-period and which are ordered over time.

time series data observations on a variable over a time-period and which are ordered over time

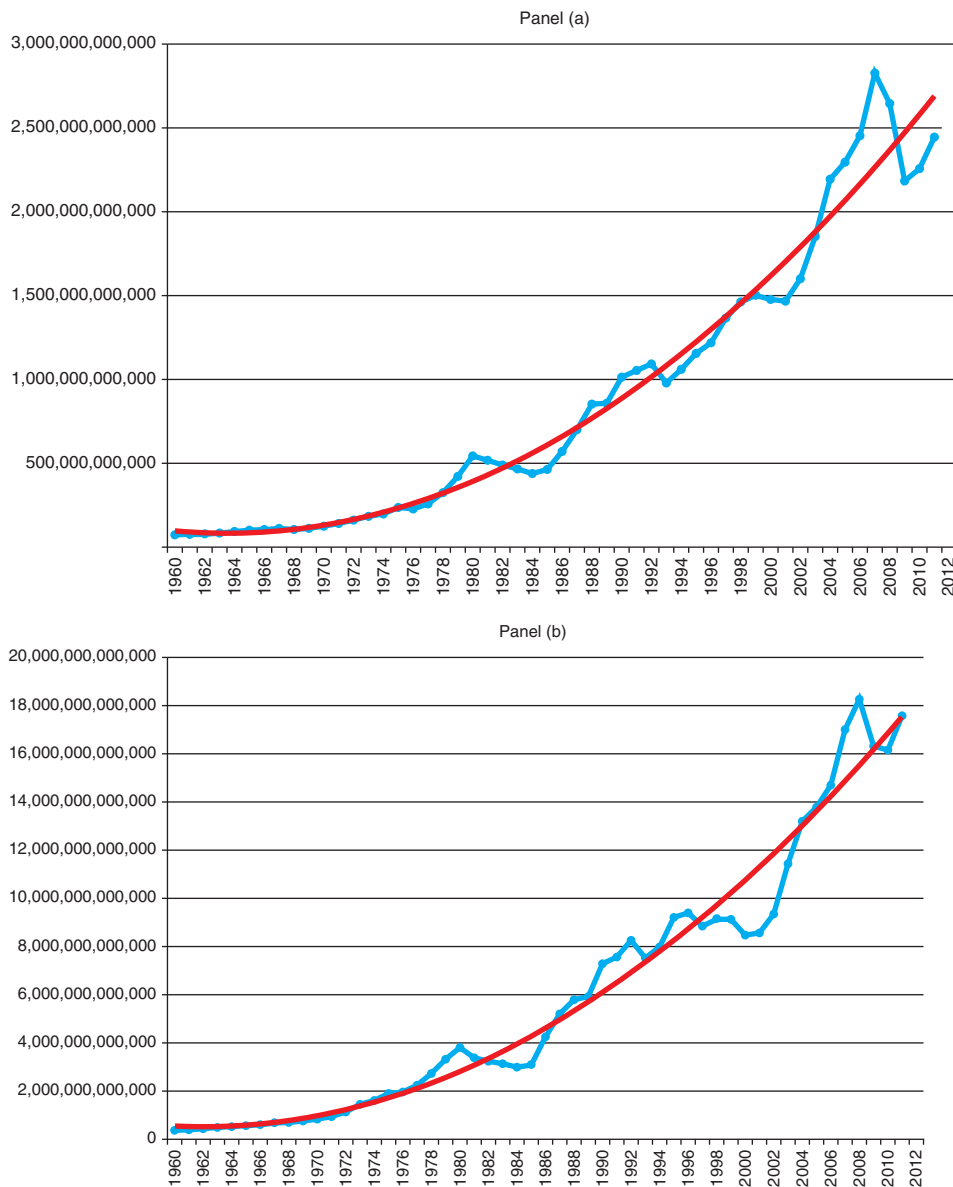
Central to the analysis of business cycles is GDP over time. Figure 30.1 shows two graphs: panel (a) shows GDP in the UK from 1960–2011. The value of GDP is given on the vertical axis in current US dollars and the horizontal axis is the time period in years. Panel (b) shows GDP across the European Union over the same time period.

Panel (a) shows that the value of the UK economy in 1960 was around \$72 billion but by 2011 the UK economy was worth around \$2.4 trillion. Panel (b) shows the economy of the EU was worth around \$365 billion in 1960 but has grown to some \$17.5 trillion in 2011. A trend line has been added in both cases and this shows how over time the GDP of both the UK and the EU has risen. It also shows how actual GDP fluctuates around the trend and that there are similarities in these fluctuations. For example, between 1960 and 1976, trend growth and actual GDP were similar but from around 1977 to the early 1980s growth was above trend but then started to decline and in the mid-1980s was below trend. This sort of pattern has been repeated with some periods above trend and some below and it is noticeable that in the period up to 2008, growth was significantly above trend but was followed by a sharp fall and growth is currently below trend. The pattern is almost the same as that for the EU as a whole.

The question arises as to what causes these fluctuations and, ever since the Great Depression in the 1930s, economists have sought to offer models to explain these macroeconomic variations. It is safe to say that there is disagreement among economists about the cause of these fluctuations and what or indeed whether any policy measures need to be put in place to deal with the welfare issues that arise. Welfare issues arise if it is assumed that the economy is deviating in some way from its equilibrium. Policy is put in place to address the welfare issues that arise. Some economists, however, believe that the economy may not be deviating from its equilibrium but instead the economy is moving from one equilibrium point to another, in which case policy is not required because there are no welfare issues arising if the economy is in equilibrium. We shall explore these differences later in the chapter.

FIGURE 30.1**GDP in the UK and Europe, 1960–2011**

Panel (a) shows UK GDP, measured in current US dollars on the vertical axis over the period 1960–2011. Panel (b) shows GDP across the European Union (also in current US dollars) over the same time period. A trend line has been added in both panels.



Data Concepts

When dealing with macroeconomic variables there are a number of concepts that we will utilize over the coming chapters which need to be understood. Figure 30.2 shows the rate of growth of UK GDP between 1960 and 2011. It is clear from this figure that the rate of growth fluctuates over the time period. From 1962, growth accelerated and reached 5 per cent in 1964 before slowing down to between 2.8 and 2.0 per cent up to 1968. Thereafter growth accelerated again reaching 7.3 per cent in 1973 before the economy slumped dramatically and economic activity actual shrank in 1974 and 1975. The pattern of rise and fall is repeated in the time period shown thereafter.

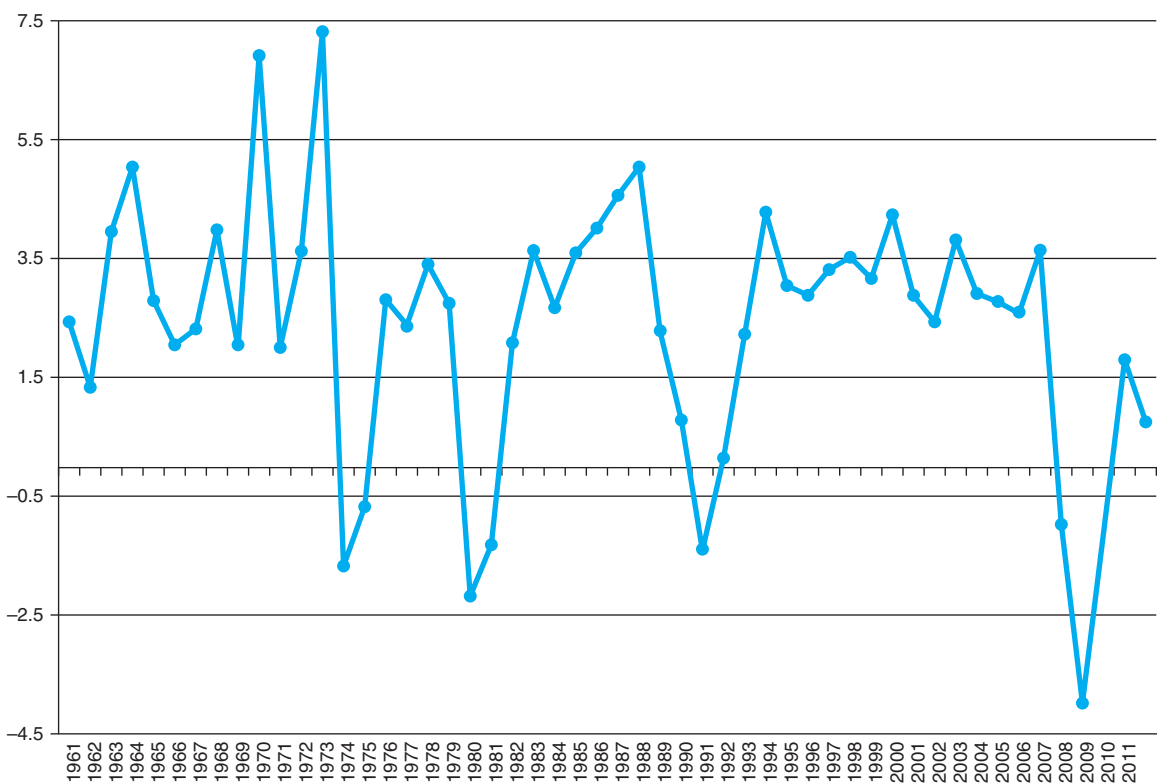
GDP shows a pattern of peaks and troughs and periods where growth is accelerating, decelerating and in some cases is declining. The **peak** is where economic activity reaches a high and real output begins to decline. A **trough** is where economic activity reaches a low and the decline ends. These are both turning points in economic activity.

peak where economic activity reaches a high and real output begins to decline
trough where economic activity reaches a low and the decline ends

FIGURE 30.2

UK GDP Growth Rate, 1960–2011 (%)

UK GDP fluctuates between various peaks and troughs over the period. In some cases, GDP growth is positive for a number of years and so the economy is growing during that time, but in other years GDP growth is negative and the economy shrinks.



When economic activity is accelerating and the rate of growth is rising each year this represents a period of expansion. Depending on the rate at which real output is accelerating, it may be referred to as a *boom*. In 2003, UK GDP grew at 3.8 per cent, in 2004 it grew at 2.9 per cent. It is important to note that economic activity slowed down in 2004 but still continued to grow. Taking a simple example to illustrate this, if economic activity is measured in terms of the number of tonnes of steel produced, if output was 10 million tonnes in 2002 and 12 million tonnes in 2003, then the growth rate is 2 per cent. If output is 13.5 million tonnes in 2004, the growth rate is now 1.25 per cent; growth has slowed compared to 2003 but the economy is still producing more steel than the year before. This is why we use the terms 'accelerating' and 'decelerating' to describe changes in the rate of growth of GDP when it is positive.

However, when real output actually shrinks then there is a **contraction** in the economy. If output of steel in 2005 was 13 million tonnes, the economy would have shrunk; less steel is being produced compared to 2004 and growth will be -3.8 per cent.

contraction when real output is lower than the previous time period

If expansion is considerably above trend then the term ‘boom’ might be appropriate but if growth is only just above trend then that term may not be applicable. It is important to be aware of the language used when discussing economic growth; the news media are keen to use emotive words such as ‘boom’, ‘bust’ and ‘slump’ when reporting growth data. Is a growth rate of 0.25 per cent ‘a slump’ for example?

In Figure 30.2 we can find the mean of the data represented by adding the sum of the annual growth rates (124) and dividing by the number of years (51). The mean growth rate over this period is 2.4 per cent. In describing periods of accelerating and decelerating growth, therefore, we can compare the growth rate at a particular time period with the mean growth rate. The period from 1970 to 1973 might reasonably be called a period of boom given that the rate of growth of GDP averaged almost 5 per cent during that time reaching a peak of 7.3 per cent in 1973. In 2008 and 2009, GDP growth was -1.0 and -4.0 per cent respectively, considerably below the mean growth rate for the whole period of 2.4 per cent.

The difference between a peak and a trough in the business cycle and trend output is called the **amplitude**. Figure 30.2 shows that this can vary considerably. In 1973 the amplitude was 4.9 but in 1984 was 0.3 and in 1981 was -3.7 . The length of the time between peaks and troughs in economic activity also varies considerably. Between 1993 and 2008, the UK enjoyed a period of persistent growth. The difference between peak and trough was 16 years. Prior to this, the difference between peak and trough (1988–1991) was just four years.

amplitude the difference between peak and trough and trend output

SELF TEST What are the main phases of the business cycle?

Trends

An important point of disagreement about economic time series data is the existence of trends. A **trend** is the underlying long-term movement in a data series. The trend can be upwards over time, downwards or constant. Earlier we took the mean (average) of the growth rate in UK GDP between 1960 and 2011 as being 2.4 per cent. The mean of a set of time series data can be used as the trend.

trend the underlying long term movement in a data series

Figure 30.1 showed UK and EU GDP between 1960 and 2011 and a trend line was added to the data showing very clearly that the trend was upwards over the time period shown. Trends can demonstrate patterns over a period which can be described as *stationary* and *nonstationary*. **Stationary data** are time series data that have a constant mean value over time, whereas **nonstationary data** are time series data where the mean value can either rise or fall over time. The data in Figure 30.1 would suggest that GDP is nonstationary data rising over time. Other economic variables might exhibit the characteristic of stationary data; unemployment, for example, tends to be more characteristic of stationary data. In the period 1993 to 2003, UK unemployment averaged 1.92 million whereas in the period from 2004–2011, unemployment averaged 1.99 million. As with any statistical analysis, care has to be taken to consider what data have been included, how the data have been constructed, what time period is involved and what test has been applied, and there is often disagreement about appropriate statistical tests and measurements used in economic analysis.

stationary data time series data that has a constant mean value over time

nonstationary data time series data where the mean value can either rise or fall over time

The disagreement about how statistical tests are used and applied is beyond the scope of this book, but understanding that disagreement between economists on vital issues, such as the path of business cycles and the policy options that might be applied, is derived in large part from different interpretations of the reliability and validity of different statistical methods. A trend over time has implications. For example, looking at Figure 30.1, would there be any reason to assume that the level of economic growth for the next 10 years will not continue to be positive and increasing in the UK and EU? Nonstationary data can be assumed to have what are called **deterministic trends** – trends that are constant, positive or negative independent of time for the series being analysed and which change by a constant amount each period. GDP might deviate from trend over short-term periods but when looked at over longer time periods will revert to the mean. In contrast, a **stochastic trend** is one where the trend variable changes by some random amount in each time period. If macroeconomic variables are analysed assuming they exhibit a deterministic trend, for example that GDP will rise by 2.4 per cent on average, then the implication is that policy measures can be applied when GDP deviates from this average. What has caused GDP to deviate from the mean must be a transitory phenomenon (is temporary and only lasting for a short period of time) and this further implies that the cause of the deviation must also be temporary and can be influenced.

deterministic trends trends that are constant, positive or negative independent of time for the series being analysed
stochastic trend where trend variables change by some random amount in each time period

What if the deterministic trend is illusory and we have applied an inappropriate statistical test to the time series data we are looking at? If the data exhibit a trend that is stochastic, then policy measures might be applied that are unnecessary and possibly distorting. If the deviations are removed the variable does not, unlike a deterministic trend, revert to the mean. Instead, the variable might move from its mean and stay away from that mean over the next time period. The law of large numbers states that, as the number of observations increases, the average of the observations is likely to be the mean of the whole population of observations. For example, if we toss a fair coin ten times, we might observe six instances of heads and four of tails. The mean of these observations in terms of the ratio heads to tails is 6:10 or 0.6. If we toss the coin a further 90 times we might observe 42 heads and 48 tails, a mean of 0.46. When we look at the mean of the population as a whole (100 observations) the mean is 0.48. If we toss the coin a further 10,000 times the chances are that the mean is likely to be close to 0.5. If we continued to toss the coin and observe the trend, it is likely that it would continue to be 0.5. Taking any selection of data from the total number of observations, the mean might be different as shown above but the greater the number of observations we observe the more likely it is that we will get the mean. It would not be unreasonable to assume that in future time periods the mean is going to continue to be 0.5.

If, however, something changes in the nature of the coin or the friction exerted by air, the surface on which the coin lands, etc., which affects the coin tossing outcomes, then the trends (mean) we observe might be completely random for every time period we observe and so when looking at the population as a whole the consistency of the law of large numbers is destroyed.

This summary of statistical arguments is important to bear in mind as we look at short-run variations in macroeconomic variables. The study of time series data is complex and it is important to take into account as we look at different schools of thought in economics, that many of the differences of opinion arise from different interpretations and use of statistics.

Procyclical and Countercyclical Movements in Macroeconomic Data

Let us assume that GDP of a country exhibits a deterministic trend and that the trend growth over a period is 2.4 per cent. If GDP over an 18-month period is reported as 1.3 per cent, what might we expect to happen to key macroeconomic variables? If GDP has slowed from trend it might be expected that the number of people out of work would have fallen, that inflation would have slowed, the money supply contracted and that real wages would have fallen. Observations over time do show that, when economic activity slows, unemployment rises and the rate at which prices rise slows. This makes intuitive sense. If firms are cutting back production, as a slowing of GDP implies, then they might not need as many

employees and so some might be let go and, at the same time, firms may think twice about increasing prices when economic activity is slowing for fear of losing sales. They may decide to keep prices constant or even lower them in an attempt to try and boost demand.

Economists will often look at the movement of pairs of variables which are called **comovements**. Typically, one of these variables is GDP. Economists will then compare another economic variable such as inflation or employment with GDP over time and see if any relationship can be determined.

comovement the movement of pairs of variables over time

Comovements may exhibit certain relationships. When a variable is above trend when GDP is above trend the variable is said to be **procyclical**. When GDP is above trend, inflation and employment also tend to be above trend and so are classed as procyclical. Real wages and nominal interest rates are also classed as procyclical. If a variable is below trend when GDP is above trend the variable is described as **countercyclical**. Unemployment can be classed as countercyclical because it tends to be below trend when GDP is above trend.

procyclical a variable that is above trend when GDP is above trend
countercyclical a variable that is below trend when GDP is above trend

SELF TEST Would you expect inflation to be procyclical or countercyclical in a period when economic activity is declining? Explain.
 What would you expect real wages to be in a period of rising economic activity? Explain.

Variables as Indicators

Economists are not fortune tellers but, in trying to understand business cycles and whether it is possible to predict changes in economic activity, they have looked at the extent to which economic variables or collections of economic variables over time give any clue to deviations from trend. Research by Wesley C. Mitchell and Arthur F. Burns in the United States in the 1930s led to an index of cyclical indicators which is now utilized by the Organization of Economic Cooperation and Development (OECD) to try and identify potential turning points in economic activity. Figure 30.3 shows the composite leading indicators index between 2003 and 2013 for the OECD area. The arrows show turning points. It is clear that the index began to decline in the middle of 2007. The global recession followed. In early 2009, the index began to rise and this indicated a turning point to suggest economic activity would begin to improve.

Cyclical indicators can have three characteristics. They can be **leading indicators** where the indicator tends to foretell future changes in economic activity, a **lagging indicator** which occurs after changes in economic activity have occurred and **coincident indicators** occur at the same time as changes in economic activity. The OECD's composite leading indicators (CLI) index is classed as a leading indicator. Figure 30.4 shows the CLI index for the UK and the Euro Area 17 from February 2011 to December 2012 (note, Latvia joined the Euro Area in January 2014). The index suggests that the UK economy might expect to see some improvement but that the situation in the Euro Area continues to be challenging.

leading indicator an indicator which can be used to foretell future changes in economic activity
lagging indicator an indicator which occurs after changes in economic activity have occurred
coincident indicator an indicator which occurs at the same time as changes in economic activity

FIGURE 30.3

OECD Area Composite Leading Indicators (CLI) Index, 2003–2013

The CLI index shows key turning points in the global economy since 2003.

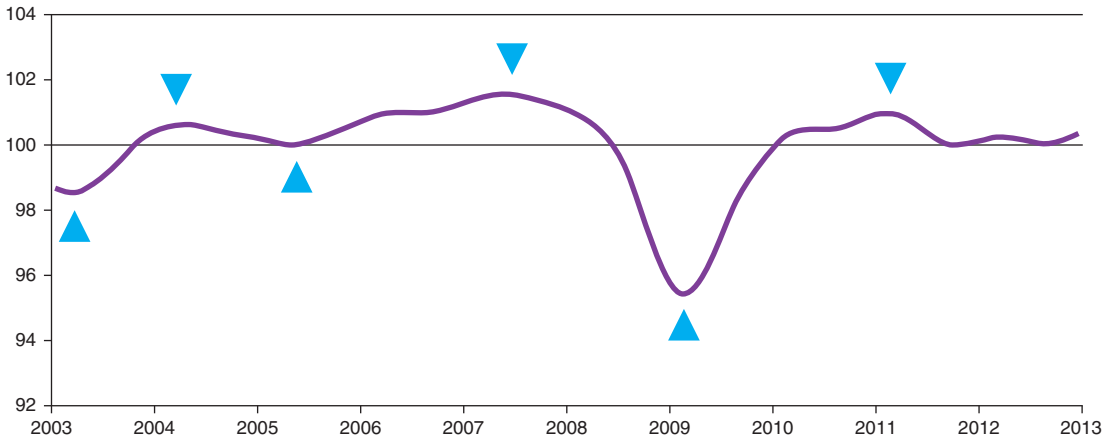
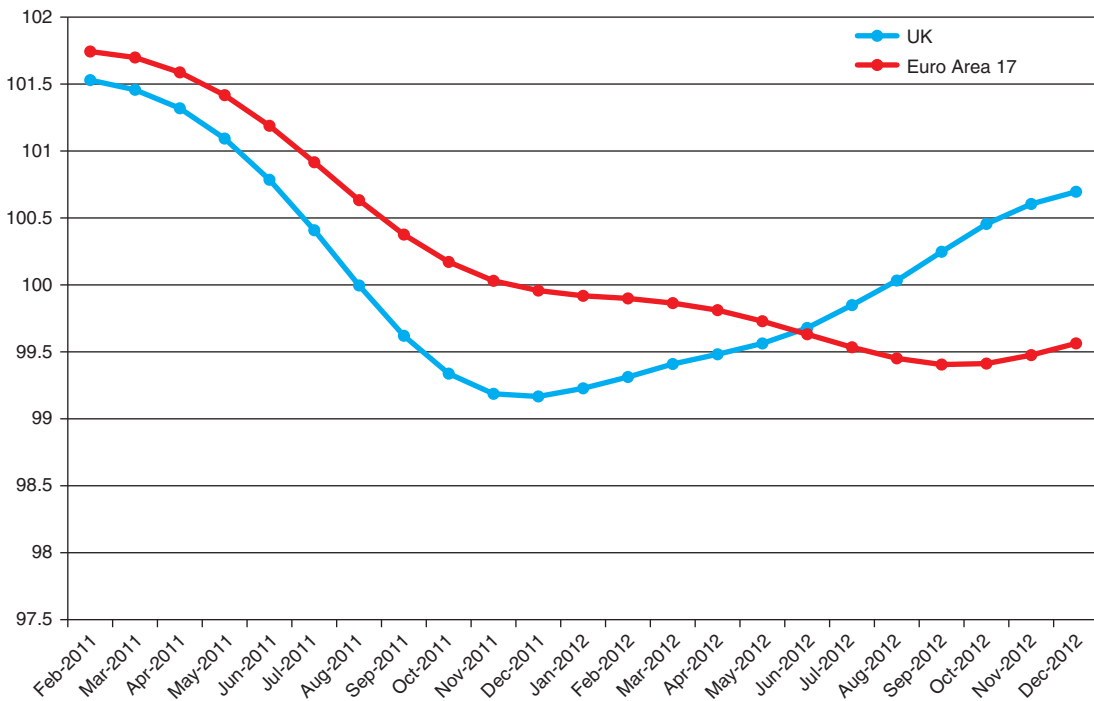


FIGURE 30.4

OECD Composite Leading Indicators Index, February 2011–December 2012 for the UK and the Euro Area 17

The upturn in the CLI index since November 2011 suggests that the UK economy is likely to improve but a softening of the index in December 2012 suggests that growth is still likely to be weak. The Euro Area 17 index has been falling for most of the period, although a slight upturn from September 2012 suggests an improvement in economic activity in the EU 17 in the future.



CAUSES OF CHANGES IN THE BUSINESS CYCLE

Having looked at some of the ways in which the data in macroeconomic time series are used and some background to the study of the business cycle, we now turn to asking questions about what causes changes in economic activity. We know that the economy is made up of households and firms and so it is reasonable to expect that the behaviour of these two elements of the economy have a role to play in how economic activity changes. Firms and households make decisions.

Household Spending Decisions

Households make decisions on how much labour to supply. The amount of labour supplied depends in part on the real wage rate. The rate of growth of wages in relation to prices affects consumers' purchasing power. How consumers perceive changes in real wages is an important factor in decision-making. Households also make consumption decisions on everyday goods and services, leisure and entertainment and also on what are sometimes called 'big-ticket' items such as durable household goods (TVs, washing machines, cars, fridge freezers etc.) not to mention decisions on purchasing houses or flats. Purchasing decisions on these items may in themselves be cyclical; if a newly-wed couple buys a house, for example, it may be that in the next year they spend considerable amounts on furniture, decorating and household goods. It then may be several years before goods begin to wear out and need replacing or the house needs decorating again.

Households will also make decisions based on changes in interest rates, house prices and taxation. Increases in interest rates may encourage a rise in saving and a reduction in consumption; changes in house prices affect people's wealth and the effect can lead to changes in spending, especially if people are able to borrow on the strength of the value of their property. Changes in tax rates affect different people in different ways and can have a considerable impact on behaviour.

Firms' Decision-Making

Firms make decisions about production levels – how much output to produce – based on what they think they can sell. If firms face strong demand they are likely to increase output and may have to take on more workers and buy more raw materials and semi-finished goods in order to satisfy demand. Some firms may look to expand by investing in new equipment and machinery or by acquiring new premises or even other firms. Firms also make decisions about how many workers to hire (or release) based on the real wage rate and productivity levels. If the real wage rate falls then firms can afford to hire more workers; if productivity levels rise then costs can be lowered and firms can be more competitive. Firms will also monitor stock levels (stock is often referred to as inventories). If stocks are building up then it may be that sales are slowing down, whereas if stocks are falling demand may be strong and sales rising. Firms will respond to changes in inventory levels by expanding or contracting output as necessary or investing in new plant and equipment or on mergers and acquisitions. The government is also a key economic actor; it makes decisions on taxes and spending and how much to borrow to finance its activities.

External Forces

Economic activity is also affected by events abroad. In an open economy, movements in exchange rates affect the competitiveness of domestic and foreign firms through changes in import and export prices. Economic activity abroad can also have an impact on countries as consumption and investment decisions by firms and consumers abroad change. The UK Chancellor of the Exchequer in 2012 regularly pointed to the problems in the euro area as a reason why the UK economy had not recovered from recession as quickly as had originally been predicted.

We also have to take into account the effect of events which are often totally unpredictable such as drought, flood, tsunamis, hurricanes, extreme cold or earthquakes, through to events such as political upheaval, war, terrorism and conflict which can have far-reaching effects on individual countries and the global economy.

Government Policy

Governments have control over a considerable amount of economic activity. Governments make decisions about tax rates and have to try and consider the incentive effects of changes in tax rates. Decisions about major infrastructure investment can be made by governments which can have ripple effects throughout an economy. Government agents can also have an effect on economic activity. Central banks are independent of most governments but have links with government. In the UK, for example, the government sets the target for inflation which the Bank of England is responsible for. The Bank of England then sets monetary policy to meet the target the government has set. Changes in interest rates will affect both households and firms through policy decisions like quantitative easing and measures designed to help ease credit flows to business and households.

Confidence and Expectations

Households and firms not only make decisions based on current needs but also on the future. It is unlikely that an individual will make a decision to take out a loan for €15,000 to buy a car if that person believes the security of their job is in question. The news media provide information on the state of the economy, governments make pronouncements, senior finance ministers and banking officials are interviewed about their forecasts for the economy. Employees will get a sense of how the firms they work for are performing and those in the public sector may take close note of government policy decisions. Workers may look at current inflation rates and their wage rise and may make judgements about inflation in the future and therefore what sort of wage rise they need to maintain their standard of living. Firms may also look at inflation rates and make judgements about price rises to consumers based on what they expect inflation to be in the future and how inflation in raw materials might affect their costs.

Our expectations of the future shape our decisions, and confidence of firms and households to make decisions is something that is very difficult to quantify and equally difficult to know when it changes. At what point does any individual begin to believe that stock or house prices have reached a peak, at what point does a firm make the decision to make redundancies and why does confidence begin to decline or to pick up? These are all difficult questions to answer but confidence and expectations play a significant part in swings in economic activity.

SELF TEST A country's finance minister presents the government's annual financial statement and advises that she expects GDP to contract sharply in the coming year because of adverse international economic conditions. How might such a speech influence confidence and expectations of households and firms?

BUSINESS CYCLE MODELS

Attempts to try and understand how business cycles occur have led to different models which differ in the assumptions that are made. On the one hand there are models which assume that markets clear quickly and as a result the welfare of economic actors (firms and households) is maximized and there are no reasons why economic actors should change their behaviour. On the other hand there are models which assume that markets do not always clear quickly and that rigidities are present in the market, particularly with regard to prices and wages which mean that for a time, economic actors will not be maximizing welfare.

We will provide an overview of the different business cycle models with perspectives that consider the supply side of the economy and then we will look at models associated with the demand side of the economy.

Supply Side – The New Classical Model

This model is based around analysis of the supply side of the economy and the operation of the labour market. It is assumed that the labour market clears but that workers have imperfect information. The model highlights the importance of anticipated and unanticipated price changes. If workers correctly anticipate

price changes they can change their behaviour such that real wages and the amount of labour they supply adjust to clear the market. For example, if the price level rises, the real wage will fall and firms will be encouraged to hire more labour. If the price change is anticipated, workers will recognize that real wages will fall and will supply less labour. The demand for labour will be greater than the supply of labour so the nominal wage will rise. However, the real wage at the new equilibrium will be constant.

If workers do not anticipate the change in prices, the real wage will fall, firms will demand more labour but workers continue to offer the same amount of labour hours so the demand for labour will be in excess of the supply and the nominal wage will rise but the rise will be less than the price rise. The result is that output will rise but the real wage will fall. The increase in output will be above trend GDP. In this model there has been a deviation from equilibrium output because workers have incomplete information and are not aware of the impact of changes in the price level.

Aggregate Supply Shocks A further aspect of the new classical model looks at how shocks to aggregate supply can cause deviations from trend output. These shocks can affect the productivity of factors of production and can be temporary such as the effect of natural disasters, or permanent such as can happen when new technologies are developed. The developments in computer technology over the last 30 years, for example, have had a permanent impact on productivity which few would have envisaged.

If supply shocks are temporary, such as the effect of an earthquake, productivity declines and demand for labour and other factors will fall which will also lead to a fall in output below trend GDP. It is possible to observe aspects of this argument in events which have occurred in the last ten years. An earthquake in Northern Japan in 2011 caused extensive damage and disruption to supply chains across the globe. Firms found that component parts were in short supply and some firms had to lay off workers and suspend production until the supply chains were re-established. The global nature of business means that such natural disasters can have far-reaching effects.

Supply Side – The Keynesian Model

In the next chapter we will look in more detail at the contribution of John Maynard Keynes to macroeconomics. One of the main propositions that Keynes put forward was that markets do not clear as quickly as classical economists believed. The Keynesian assumption is that the ability of the goods and labour market to clear are impaired by the existence of *sticky prices* and *sticky wages*.

In labour markets, firms enter into contractual agreements with workers and are also constrained by labour market legislation and regulation which means that it is not always easy to adjust the labour force to changed economic circumstances. Excess demand or supply in the labour market will not be eliminated quickly by changes in wages because of these wage rigidities, and in particular wages tend to be sticky downwards in that it is difficult for firms to adjust wages down when there is an excess supply of labour as a result of changing economic conditions, such as a change in the price level.

Sticky prices occur where there are costs to firms of changing prices. We introduced the idea of menu costs in Chapter 27. Changing economic conditions in the goods market may warrant a change in price to clear the market but because firms face costs in changing prices it may be that prices are changed infrequently. In addition to the internal costs to firms of diverting labour resources to changing prices, firms will have agreements with suppliers and retailers about prices built into contracts which may not be capable of being re-negotiated for some time. The existence of menu costs mean that prices will be sticky and prevent markets from clearing in the short run.

Demand Side – The New Classical Model

We have noted in earlier chapters that aggregate demand is composed of consumption spending, investment spending, government spending and net exports. Changes to any or all of the components of aggregate demand could cause a deviation of output from trend. Later in the book we will look at causes of shifts in aggregate demand but in this section we are going to trace through the effect of an assumed change in aggregate demand. A rise in aggregate demand, for example, will, *ceteris paribus*, lead to a rise in the price level. A rise in the price level reduces real wages and firms look to hire more workers with the result that the nominal wage will rise. The rise in aggregate demand will lead to an increase in output and the

price level increases. The increase in the number of workers hired causes a fall in unemployment. The New Classical interpretation rests on workers misinterpreting a rise in nominal wages as a rise in real wages, in other words, they do not fully take into account the effect on wages of the price rise. We referred to this in Chapter 27 as the *inflation fallacy*.

The result of this is that the economy moves to a temporary equilibrium where the expectations of some economic actors are not fully incorporated because they are incorrect. However, over time, workers will begin to realize that real wages have changed and as a result begin to change their behaviour. As workers negotiate for wage rises that maintain their standard of living, firms' costs rise and some firms will cut back supply with the result that the economy returns to trend output but with a higher price level once expectations have fully adjusted.

If aggregate demand falls then the economy will enter a period of contraction with output and prices falling in the short run. The reverse process to the one outlined above will take place. Real wages rise and firms begin to cut back on output which increases unemployment. The demand for labour will be in excess of the supply of labour and nominal wage rates will fall. Workers see the fall in nominal wages as a fall in the standard of living but over time their expectations will adjust to take into account the fall in the price level and output will return to trend.

Cyclical Implications The analysis provided above has implications for the nature of the cyclicity of key macroeconomic variables. In the new classical model, when output is above trend, unemployment is countercyclical and employment will be above trend and so be procyclical. Inflation will be procyclical but real wages will be countercyclical because as output rises real wages fall.

Demand Side – The Keynesian Model

Remember that the Keynesian model assumes sticky prices and wages. If there is an increase in aggregate demand then wages and prices will take time to adjust. The increase in demand will mean firms' stocks begin to decline and so they will take steps to increase output and in so doing increase employment. In the short run, therefore, output increases above trend but the price level does not change because of sticky prices.

Over time, however, the economy will return to trend because firms will eventually be able to raise prices and nominal wages will also increase. The rise in nominal wages affects firms' costs and some will begin to cut back output which returns to trend but with a higher price level (which is the same outcome as that given in the new classical model above), but the way in which the economy has adjusted to the deviation from trend has differed. If aggregate demand is reduced, the reverse situation will apply and explains why recessions occur.

The speed with which the economy returns to trend after an aggregate demand shock will depend on the time it takes for prices and wages to adjust to the changed economic conditions. In this model, employment, real wages and inflation are procyclical and unemployment is countercyclical.

Real Business Cycles

At the heart of the model of real business cycles is the belief that changes in technology, both positive and negative technology shocks, affect productivity regardless of the real wage rate. The model assumes that there are no market imperfections, that firms and households are profit and utility maximizing and that markets clear. Against the background of these assumptions, if there is a negative technology shock then labour productivity falls and the demand for labour falls. Output will fall as a result and unemployment will increase. Output falls because aggregate supply falls, which creates excess demand in the economy. An excess of aggregate demand will lead to a rise in the price level. If the price level increases this affects the real interest rate. Recall that the real interest rate is the nominal interest rate minus inflation. If inflation rises and the nominal interest rate is constant then the real interest rate will rise. A rise in the real interest rate would lead to a fall in investment by firms.

The causes of business cycles, therefore, are technology shocks that cause permanent shifts in aggregate supply, but when aggregate supply shifts the expectations of economic actors are still correct and so there is no reason for governments or central banks to intervene and apply policy prescriptions. The real business cycle model, therefore, does not see growth over time as being a deterministic trend but a stochastic one. In real business cycle models, employment, labour productivity and real wages are procyclical.

CASE STUDY

Kydland and Prescott and Real Business Cycles

Finn E. Kydland and Edward C. Prescott won the Nobel Prize for Economics in 2004 and their award was based on the work they had done on business cycles. Their work became associated with real business cycles because of the focus on real shocks to the economy as opposed to nominal shocks as the key driver of deviations in output. The question Kydland and Prescott were interested in was not measuring business cycles but focusing on the *pattern* of output and employment around the trend and asking why this seemed to happen in a repeated fashion over time.

Many textbooks explain the business cycle in terms of four phases – boom, slowdown, recession, upturn. The term ‘cycle’ implies a ‘what goes around comes around’ type of approach to an economy. Such a view of cycles implies an almost inevitable trend where growth turns into boom, which in turn leads to the start of decline that leads to recession before the process begins again. However, Kydland and Prescott used a generally agreed scientific definition of ‘cycles’ that makes reference to a point of departure – in this case the trend of economic growth. Kydland and Prescott refer to these recurrent departures as ‘deviations’.

They argue that business cycles must be seen as periodic deviations from trend growth and that business cycles are neither inevitable nor evolutionary. As such the explanation for a downturn in economic activity cannot, in itself, be found in the reasons why growth occurred in the first place. Equally, the seeds of an expansion in growth are not to be found in a recession.

In analysing the behaviour of these deviations from trend they sought to challenge some conventional wisdom that had grown up around business cycles. For example, in times of economic downturn the expectation would be that the price level would fall and in times of strong economic growth it would be anticipated that prices would rise – in other words, inflation is procyclical. This implies that in a downturn firms seek to reduce prices to encourage sales and are prevented from increasing prices to improve margins because of the lack of demand. Similarly, in times of economic growth, firms experience rising demand and possibly wage and other costs. They are able to increase prices to improve margins without too much damage to business because of the strong growth in demand.

Kydland and Prescott argued that, in fact, price showed a countercyclical behaviour – when economic growth slowed down, prices rose and when economic growth was strong, prices fell. They further argued that real wages fall as growth increases and vice versa (or are not related to the business cycle) and that the money supply was an important factor in leading economic growth.

Price procyclicality is important because if we are looking for the causes of changes in economic activity as a whole, we would presume to be looking for something fairly major as being the cause – large price rises, for example, or shocks caused by changes in things like the money supply. If price procyclicality is a myth then research into the causes of changes in cycles might be misguided. Think of it in terms of a thermometer in a room. The thermometer tells us what the temperature of the room is but is not a cause of the temperature. Looking at the properties of the thermometer to explain the temperature of the room would lead us down the wrong path.

Kydland and Prescott argue that an important factor in explaining business cycles is the decisions people make about how they devote their time between leisure (non-market activities) and income-earning activities. After analysing the factors that may influence business cycles they come to the following conclusions:

- Aggregate hours worked (a measure of labour input) is strongly correlated with changes in GDP. The problem with this is that the contributions to GDP for all workers are considered the same. Kydland and Prescott point out that the contribution by the hours worked by a brain surgeon is not the same as that of a porter in a hospital. With some consideration of this, Kydland and Prescott conclude that real wages are more procyclical and something which traditional literature on business cycles would not suggest.
- The capital stock is largely unrelated to real GDP but is closely correlated if a time lag of about one year is included.
- With regard to the factors affecting aggregate demand – consumption, investment and government spending – Kydland and Prescott report that consumption and investment are highly procyclical whereas government spending does not seem to be correlated with growth.

(Continued)

- They also comment that imports are procyclical as are exports but with a six-month to a year time lag.
- Labour and capital income is strongly procyclical.
- They find no evidence that narrow money (M1) leads the business cycle. In other words, they do not find evidence that a rise in M1 will lead to a spurt in growth.
- Credit arrangements are likely to play a significant role in future analysis of business cycle theory.
- The price level is countercyclical.

Kydland and Prescott's work has prompted considerable research not least the necessity to look at what might be happening in the macroeconomy. It might be necessary to look at factors other than those that simply describe the data. This emphasis on the quantitative features rather than qualitative features (what it tells us rather than what we think it might signify) has been a feature of a re-assessment of statistical analysis in economics, particularly the analysis of time series data. What do long-term time series tell us in relation to short-term series?

Real business cycle theory does not view a recession as a 'failure' in the economy nor might a boom also be seen as a failure. (A boom might be interpreted as a failure in economic management because it is unsustainable which is why politicians often refer to the 'bad old days of boom and bust'.) Kydland and Prescott see business cycles as explanations of shock to the economy that are understandable reactions rather than failures. Their work tends to dismiss the 'sticky prices' explanation for a slowdown in growth and also the mismatch between investment and consumption and the monetarist argument of market failure in price signals. Instead they look at real shocks to the economy and the adjustment process to those shocks, which could last for some time after. Essentially, Kydland and Prescott argued that business cycles could occur perfectly naturally within a competitive environment despite the implication in traditional theory that, for example, perfect competition would not result in long periods of unemployment.

Real business cycle theory is not without its critics. In particular, the implication that shocks to supply tend to be permanent rather than transitory has been questioned as an explanation of recessions, and empirical studies have also questioned the extent to which the assumptions of the model and the predictions about the cyclicity of economic variables match the evidence.



Finn E. Kydland (left) and Edward C. Prescott (right); winners of the Nobel Prize for Economics in 2004.

CONCLUSION

In this chapter we have provided some background to the more extensive analysis of short-term economic fluctuations in the coming chapters. We have introduced a number of key concepts which need to be borne in mind as we tackle the next chapters, not least building a familiarity with the behaviour of key macroeconomic variables over time.

We have also learned that economists have different interpretations of business cycles based on different assumptions about how the economy works and the extent to which markets clear and how economic actors behave. Models of business cycles focus on both the supply side and demand side of the economy and it is likely that the answers to the questions surrounding policy decisions to help smooth out deviations in trend growth are to be found in a combination of these different models. The model of the real business cycle raises questions about whether the trends that we 'see' in time series data, which form the basis of analysis, are actually present and if trends do not exist then policy measures designed to reduce deviations from trend are misguided.

IN THE NEWS

Employment and Unemployment in an Economic Downturn

We have seen that key macroeconomic variables such as employment and unemployment behave in particular ways during periods of deviation of output from trend. However, the experience of the UK economy in recent years has caused economists to re-assess their assumptions about cyclicalities and seek explanations for observed events.

Weak Growth and Rising Employment

In the time period between the three months to December 2011 and the same period in 2012 the number of people working full-time in the UK rose by 394,000 and the number of people working part-time rose by 190,000. This represented the largest increase since 2005 and meant that the number of people employed in the UK was recorded as 29.73 million or an employment rate of 71.5 per cent of the population aged 16–64. The number of people unemployed in the three months to December 2011 stood at 2.66 million and in the same period in 2012 the figure was 2.51 million. The claimant count in January 2012 stood at 1.60 million and in December 2012 at 1.55 million.

In themselves, these figures might not seem particularly remarkable but when set against the performance of the economy during a similar time period they do raise

some interesting questions which have led economists to look closer at the data to find answers. Gross domestic product (GDP) during this period has been sluggish at best. During 2011, growth was reported at under 1.0 per cent for three of the four quarters, at just 0.4 per cent in the first quarter of 2012, –0.8 per cent in the second quarter, a bounce back due to the Olympic Games effect in the third quarter of 1.6 per cent and a 0.1 per cent rise in the fourth quarter. The Office for National Statistics reported that GDP in volume terms shrank by 0.3 per cent in the fourth quarter of 2012 and Her Majesty's Treasury (HMT) reported real GDP in 2010 as £1,427 billion in 2010 and £1,440 billion in 2011.

In a time, therefore, when the economy of the UK is performing poorly, it might be expected that if employment is procyclical and unemployment countercyclical, why are both behaving contrary to expectations? For some time, economists and analysts have been asked to explain

what is happening. The answers might lie in a number of areas.

Employers know that making staff redundant would help to cut costs in the short run but they might also be looking at the type of employees they have. If these employees have important skills it might not be cost-effective to lose them only to have to suffer the cost of having to recruit staff when the upturn comes and possibly train up staff to the levels of those they have already. As a result, some firms may be playing a waiting game and making a judgement that short-term benefits in terms of cost saving might be outweighed by the medium-term disadvantages of not having the right staff at the right time and so are holding onto employees.

It might also be the case that employees are taking into account the prospects facing them in a tough economic environment and agreeing far more flexible working conditions with employers in an effort to hold onto their jobs. This might

(Continued)

include accepting reduced working hours or agreeing cuts in pay to help employers to continue to afford to keep workers on. Average earnings growth was under 2 per cent in 2011 and slowed from 1.7 per cent to 1.4 per cent in 2012. With inflation running above 2.5 per cent over these two years, real wages fell which may also help to explain why employment rose or why unemployment did not rise as much as might be expected in an economic downturn.

The number of people accepting part-time work or being employed in jobs for which they are over-skilled (a situation termed under-employment) has also risen. Those getting part-time jobs might prefer full time work but are willing to take part-time jobs rather than have to claim benefits and might also believe that having some job puts them in a better position to apply for full time work when it becomes

available. There has also been an increase in the number of people becoming self-employed with some 367,000 registering as self-employed between 2008 and 2012 according to the ONS. If people are working part-time or are self-employed then they are not unemployed even though they might not be in a 'first choice' employment position.

The Institute for Fiscal Studies (IFS) in the UK published a report in February 2013 which suggested that there was little evidence for the idea that firms were 'hoarding' staff in the expectation of an upturn in the economy. Whilst employment levels have remained 'robust' and the unemployment 'mercifully low', contrary to other periods of recession experienced in the UK, it is productivity which has suffered. The IFS suggests that productivity per hour per worker declined by 2.6 per cent in 2013 compared to the beginning of

2008. What this means is that the UK is producing 12.8 per cent less than it would have been if pre-recession levels of productivity growth had been maintained. The reasons why employment has been rising and unemployment rising by less than expected is that real wages are low, business investment has been weak (16 per cent lower than its pre-recession peak) and capital is misallocated. Capital is misallocated because banks are being more sympathetic to firms who might be in financial trouble or who are suffering from lower productivity and are also more risk averse in their investment strategies which together reduce the entry and exit of firms from industry.

Questions

- 1 What is meant by the terms 'procyclical' and 'countercyclical'.
- 2 Why would you expect employment to fall and unemployment to rise during an economic downturn?
- 3 What economic arguments are there for firms to 'hoard' workers in anticipation of better economic times to come? How persuasive do you find these arguments?
- 4 What effect do falling real wages have on firms' decisions on the demand for labour?
- 5 The IFS report noted that whilst both employment and unemployment had not been affected in the same way as in previous UK recessions, the cost has been a sharp fall in productivity. What might be the consequences of the fall in productivity for the UK economy?

Productivity is a measure of output per worker per period of time. Unemployment may not be rising as fast in the UK as expected in a recession but how productive are those that are employed?



SUMMARY

- Economies experience periods of changing levels of economic activity.
- Key macroeconomic data are often time series data which raises questions about the validity and reliability of statistical tests applied to data sets.
- Economic growth appears to follow a trend which rises over a period of time.
- Deviations from this trend are known as the business cycle.
- The business cycle has characteristic features which include peaks in economic activity, slowdown, troughs and upturns with key turning points.
- Trends can be stationary and nonstationary. Non-stationary data can exhibit deterministic trends which change by a constant amount independent of time or be stochastic where the trend variable changes by a random amount.
- Macroeconomic data can be viewed in pairs with one of the pairs generally GDP. The variable compared may be either procyclical or countercyclical.
- Collecting macroeconomic data can allow economists to view certain indicators as leading, coincident or lagging.
- Changes in the business cycle can be caused by changes in decision-making by households and firms, by external shocks, government policy and changes in confidence or expectations of the future.
- Business cycle models differ in the assumptions they make about the extent to which markets clear and the relationship between the supply side and demand side of the economy.
- Real business cycles emphasize the effects of changes in technology as causes of changes in economic activity.

QUESTIONS FOR REVIEW

- 1 What is time series data? Give three examples of key macroeconomic variables that are examples of time series data.
- 2 What are the main stages of a typical business cycle and how long do these stages last?
- 3 What is the difference between a deterministic trend and a stochastic trend?
- 4 Would you expect the following variables to be procyclical or countercyclical if GDP was above trend? Explain.
 - a. inflation
 - b. unemployment
 - c. employment
 - d. real wages
 - e. nominal interest rates.
- 5 Unemployment is classified as a lagging indicator. Explain what this means and why unemployment is classed as such.
- 6 Why might changes in household decision-making cause a deviation in GDP from trend?
- 7 What role does household and firms' confidence play in business cycles?
- 8 What is the role of unanticipated price changes in the new classical model of business cycles?
- 9 Why do Keynesian models of the business cycle emphasize that markets do not always clear immediately?
- 10 What is the key difference between real business cycle models and other business cycle models?

PROBLEMS AND APPLICATIONS

- 1 Look at Figure 30.1. For either the UK or the EU:
 - a. estimate the amplitude of the deviations from trend over the period shown
 - b. estimate the length of time between the beginning of the deviation from trend and the return to trend in each case.
- 2 In one time period the growth rate of GDP of a country is recorded as 3.4 per cent against a trend growth rate of output of 2.8 per cent. In the next two time periods, the growth rate of GDP is recorded as 2.8 per cent and 2.0 per cent respectively. Is the country experiencing a recession? Explain.
- 3 Go to the national statistics office website for your country (this could be the Office for National Statistics in the UK or Eurostat for the EU) and look up unemployment and inflation figures over the last 30 years. Plot the data on a spreadsheet, graph it and identify the trend. Is the trend stationary, nonstationary or indeterminate? Explain.

- 4 Is it always possible to discern a trend in any time series data? What problems might arise if trends are apportioned to time series data that are not really present?
- 5 Look up the following comovements for the country in which you are studying on an appropriate website. Explain whether the comovements are procyclical or countercyclical:
 - a. GDP and inflation
 - b. GDP and employment
 - c. GDP and unemployment
 - d. GDP and the money supply (M1).
- 6 Firms experience a rise in stocks. Explain why this might have occurred and what you expect firms' response to this event might be and how this might affect output.
- 7 Why might household and firms' confidence and expectations change leading to deviations in output from trend? Is there any way in which changes in confidence can be measured to provide an indicator of changes in economic activity?
- 8 There is a fall in the price level which workers do not anticipate. Explain what effect such a scenario will have on output.
- 9 To what extent do you think that workers are always fooled by the inflation fallacy?
- 10 Which business cycle model do you find the most compelling and why?

31

KEYNESIAN ECONOMICS AND IS-LM ANALYSIS

The next four chapters will cover different aspects of macroeconomic policy, in particular the role of fiscal and monetary policy. The framework for analysing the effects of these two policies is developed through a series of steps leading to the model of aggregate demand and aggregate supply.

In 1936, economist John Maynard Keynes published a book entitled *The General Theory of Employment, Interest and Money*, which attempted to explain short-run economic fluctuations in general and the Great Depression in particular. In the last chapter we looked at models which explained the business cycle. Prior to *The General Theory*, the classical assumption of market clearing was prevalent and it can be argued that the development of macroeconomics as a separate path of economic research stemmed from the experiences of the Great Depression and a desire to try to understand why it had appeared that the classical model had gone so spectacularly wrong. Keynes was not the first to question classical paradigms; in the preceding chapter the work of Wesley C. Mitchell in the United States was alluded to. Mitchell published research on business cycles some 20 years before the Great Depression. Keynes picked up this tradition and helped create interest in the analysis of aggregate phenomena.

Keynes' primary message was that recessions and depressions can occur because of inadequate aggregate demand for goods and services. Keynes had long been a critic of classical economic theory because it could explain only the long-run effects of policies. A few years before offering *The General Theory*, Keynes had written the following about classical economics: 'The long run is a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is long past, the ocean will be flat.'

Keynes' message was aimed at policymakers as well as economists. As the world's economies suffered with high unemployment, Keynes advocated policies to increase aggregate demand, including government spending on public works. Keynes argued for the necessity of short-run interventions in the economy. He argued that such intervention could lead to improvements in the economy that would be beneficial rather than waiting for the long run equilibrium to establish itself – hence the famous quote 'In the long run we are all dead'.

The focus on monetary and supply-side policy as the main ways of controlling the economy in most developed countries in Europe had largely consigned Keynesian demand management to the economic history books. However, the financial crisis and subsequent recession of 2007–2009 has reignited the debate about the role of Keynesian economics in macro policy. Keynes' contribution to economic thinking is widely acknowledged and it is valuable to have some insight into Keynesian economics. This chapter will begin this process.

THE KEYNESIAN CROSS

Classical economics placed a fundamental reliance on the efficiency of markets and the assumption that they would clear. At a macro level, this meant that if the economy was in disequilibrium and unemployment existed, wages and prices would adjust to bring the economy back into equilibrium at full employment. **Full employment** is defined as a point where those people who want to work at the going market wage

level are able to find a job. Any unemployment that did exist would be classed as voluntary unemployment. The experience of the Great Depression of the 1930s brought the classical assumptions under closer scrutiny; the many millions suffering from unemployment could not all be volunteering to not take jobs at the going wage rates so some must, therefore, be involuntarily unemployed. We have also seen in the preceding chapter that prices can also be sticky, meaning markets may not always adjust to clear surpluses and shortages quickly.

full employment a point where those people who want to work at the going market wage level are able to find a job

Planned and Actual Spending

Fundamental to Keynesian analysis is the distinction between *planned* and *actual* decisions by households and firms. **Planned spending, saving or investment** refers to the desired or intended actions of firms and households. A publisher may plan to sell 100,000 copies of a textbook in the first three months of the year; an individual may plan to go on holiday to Turkey in the summer and save up to finance the trip; a person may intend to save €1,000 over the year to put towards paying for a wedding next year.

planned spending, saving or investment the desired or intended actions of households and firms

Actual spending, saving or investment refers to the realized, *ex post* (after the event) outcome. The publisher may only sell 80,000 copies in the first three months and so has a build-up of stock (inventories) of 20,000 more than planned; the holidaymaker may fall ill and is unable to go on holiday and so their actual consumption is lower than planned (whereas actual saving is more than planned because they have not spent what they intended) and the plans for saving for the wedding may be compromised by the need to spend money on repairing a house damaged by a flood.

actual spending, saving or investment the realized or *ex post* outcome resulting from actions of households and firms

Planned and actual outcomes might be very different as briefly outlined above. As a result Keynes argued that there was no reason why equilibrium national income would coincide with full employment output. Wages and prices might not adjust in the short run because of sticky wages and prices, and so the economy could be at a position where the level of demand in the economy was insufficient to bring about full employment. The mass unemployment of the 1930s could be alleviated, he argued, by governments intervening in the economy to manage demand to achieve the desired level of employment.

The Equilibrium of the Economy

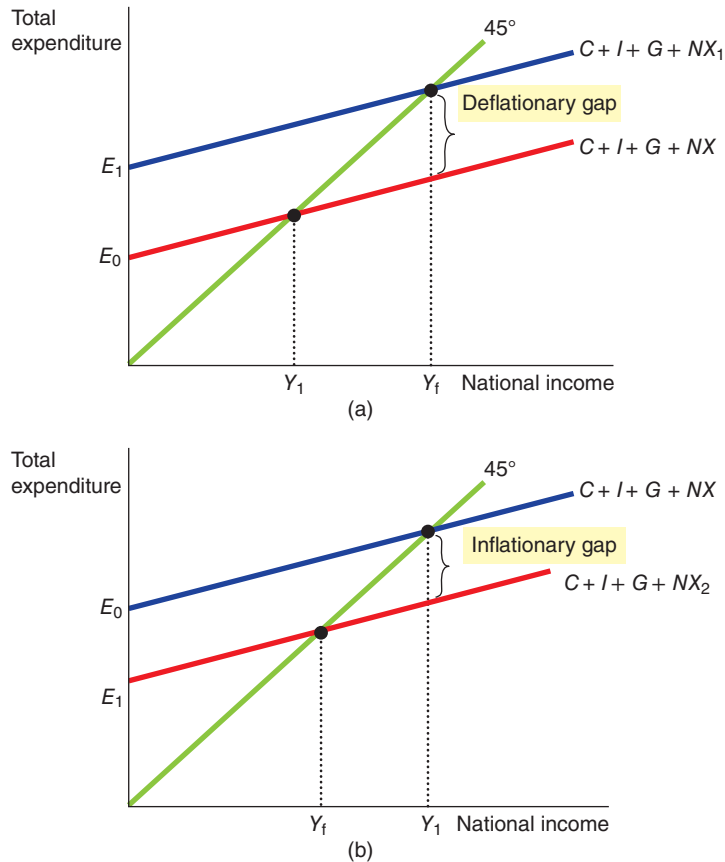
Recall the identity describing how a country's gross domestic product (which we will refer to as national income, Y) is divided among four components: consumption spending, investment spending, spending by government and net exports – the difference between the funds received from selling exports minus the expenditure on imports.

Imagine a situation where, at every point, total expenditure in the economy given by $C + I + G + NX$ was exactly the same as national income. We could represent this in diagrammatic form as a 45° line such as the one in Figure 31.1.

FIGURE 31.1

Deflationary and Inflationary Gaps

The 45° line shows all the points where consumption spending equals income. The vertical intercept of the expenditure line shows autonomous expenditure. The economy is in equilibrium where the expenditure line, $C + I + G + NX$ cuts the 45° line. In panel (a) this equilibrium is lower than full employment output (Y_f) at Y_1 – there is insufficient demand to maintain full employment output. The government would need to shift the expenditure line up to $C + I + G + NX_1$ to eliminate the deflationary gap as shown. In panel (b) the equilibrium is higher than full employment output – the economy does not have the capacity to meet demand. In this case the government needs to shift the $C + I + G + NX$ line down to $C + I + G + NX_2$ to eliminate the inflationary gap.



In panels (a) and (b), the 45° line connects all points where consumption spending (actual expenditure) would be equal to national income (planned expenditure). This line can be thought of as the equivalent of the capacity of the economy – the aggregate supply (AS) curve.

The $C + I + G + NX$ line is a function of income – in other words, spending depends on income. If income is higher, spending will also be higher and so the $C + I + G + NX$ line has a positive slope. The vertical intercept of the $C + I + G + NX$ line given as E_0 , is termed **autonomous spending** or **autonomous expenditure**. This is the component of expenditure which does not depend on income/output – government spending being a key element of this expenditure.

autonomous spending or **autonomous expenditure** spending which is not dependent on income/output

Where actual spending is equal to planned spending is the short-run equilibrium of the economy. Note that the use of the term 'equilibrium' in this context does not mean the 'best' or 'desired' equilibrium – it is simply a point where actual spending is equal to planned spending. The economy is in equilibrium where the $C + I + G + NX$ line cuts the 45° line. This is referred to as the *Keynesian cross*. In panel (a) the economy is in equilibrium at a national income of Y_1 . However, full employment national income is at Y_f . Actual spending of $C + I + G + NX$ in panel (a) gives an equilibrium which is less than that required for full employment output (Y_f). At the equilibrium Y_1 there is spare capacity in the economy – some resources are not being used to their full extent, capital may be underused and unemployment will exist. This is the equivalent to an economy being at a point inside its production possibilities frontier which we looked at in Chapter 19. The difference between full employment output and the expenditure required to meet it is termed the **deflationary gap** (you may sometimes see this also referred to as the **output gap**). Expenditure needs to rise to $C + I + G + NX_1$ in order to eliminate the deflationary gap which is the vertical distance between actual spending and spending necessary to achieve full employment.

deflationary gap or **output gap** the difference between full employment output and expenditure when expenditure is less than full employment output

In panel (b) the $C + I + G + NX$ line cuts the 45° line at an output level Y_1 which is higher than full employment output, Y_f . In this situation the economy does not have the capacity to meet actual spending. This will trigger inflationary pressures in the economy. The difference between full employment output and the expenditure line here is called the **inflationary gap**. Actual spending needs to be reduced to eradicate the inflationary gap and the $C + I + G + NX$ line needs to be reduced to $C + I + G + NX_2$ to bring the economy to an equilibrium where actual spending equals full employment output.

inflationary gap the difference between full employment output and actual expenditure when actual expenditure is greater than full employment output

Demand Management The deviations in the business cycle in Keynesian analysis are primarily due to demand side factors. The principle behind Keynesian economics is at its heart: very simple and intuitive. Downturns in economic activity occur because firms fail to sell all the goods and services they planned to sell. If customers (and of course we are not only talking about final consumers, we are also talking about other businesses as customers of firms) are not buying as many goods and services, firms will not need to produce as many and so cut back production as stocks rise. If production is cut back then firms do not need as many workers and either do not replace workers when they retire, make some workers redundant or lay-off workers by reverting to shorter working weeks or even ceasing production temporarily for a time period. Unemployment rises and the cause is due to *demand deficiency*. The affected workers now see a fall in their incomes and so cut back on spending which exacerbates the problem.

The cause of a fall in consumption in the first place is often difficult to pinpoint as noted in the preceding chapter. It could be due to confidence and expectations; it could be due to the way in which the public respond to news items (the more the news media refer to the possibility of slowdown or recession it might become a self-fulfilling prophecy); it could be due to a change in patterns of consumption with some firms seeing a decline in demand for their goods whilst others see an increase but the structural change causes a disruption in demand. Workers who are made redundant from the declining industries may not have the skills to move to jobs being created in the growth industries and as a result, regardless of the wage rate, they remain unemployed.

Keynes argued that governments can use the tools of fiscal and monetary policy, and in particular fiscal policy, to influence demand in the economy and reduce deflationary and inflationary gaps. Recall in

Chapter 20 we introduced a more detailed circular flow of income diagram which included leakages and injections. Taxation is a leakage from the circular flow but can be manipulated by the government. Equally, government can vary its own expenditure and the combinations of changes in tax and government spending can be used as levers to manage demand to bring the economy into equilibrium at a point nearer to full employment output. If the value of full employment output, the amount the economy is capable of producing if all existing resources are used fully (planned spending) is €1 trillion, (€1,000,000,000,000) for example, but actual spending is only €800 billion (€800,000,000,000). The deflationary or output gap would be €200 billion. The government might introduce policy levers which lead to a cut in taxes and a boost to government spending to generate this additional €200 billion in spending. The use of these 'levers' has some interesting features which we will describe in the next section.

SELF TEST Why might actual spending differ from planned spending?

If planned spending in an economy is €500 billion but actual spending €400 billion, is there an inflationary gap or a deflationary gap? Explain.

What might the government do in this situation to bring spending more in line with full employment output?

THE MULTIPLIER EFFECT

The $C + I + G + NX$ line is referred to as the expenditure function. Planned expenditure (E) is dependent on the level of consumption, plus investment, plus government spending plus net exports and can be written:

$$E = C + I + G + NX$$

Actual expenditure/output (remember that expenditure is one way of measuring output – these two things are the same) will be denoted as (Y). The economy will be in equilibrium, therefore, when planned expenditure is equal to actual expenditure ($E = Y$).

The positive slope of the expenditure function implies that planned spending rises as income rises. What determines the slope of the expenditure function? This is what we will explore in this next section.

When a government makes a purchase, say it enters into a contract for €10 billion to build three new nuclear power generating stations, that purchase has repercussions. The immediate impact of the higher demand from the government is to raise employment and profits at the construction company (which we shall call Nucelec). Nucelec, in turn, has to buy resources from other contractors to carry out the job and so these suppliers also experience an increase in orders. Then, as the workers see higher earnings and the firm owners see higher profits, they respond to this increase in income by raising their own spending on consumer goods. As a result, the government purchase from Nucelec raises the demand for the products of many other firms in the economy. Because each euro spent by the government can raise the aggregate demand for goods and services by more than a euro, government purchases are said to have a **multiplier effect** on aggregate demand.

multiplier effect the additional shifts in aggregate demand that result when expansionary fiscal policy increases income and thereby increases consumer spending

The multiplier effect continues even after this first round. When consumer spending rises, the firms that produce these consumer goods hire more people and experience higher profits. Higher earnings and profits stimulate consumer spending once again, and so on. Thus, there is positive feedback as higher demand leads to higher income, which in turn leads to even higher demand. Once all these effects are added together, the total impact on the quantity of goods and services demanded can be much larger than the initial impulse from higher government spending.

The multiplier effect arising from the response of consumer spending can be strengthened by the response of investment to higher levels of demand. For instance, Nucelec might respond to the higher demand for building services by buying more cranes and other mechanized building equipment. In this case, higher government demand spurs higher demand for investment goods. This positive feedback from demand to investment is sometimes called the *investment accelerator*.

CASE STUDY The Accelerator Principle

The accelerator principle relates the *rate of change* of aggregate demand to the *rate of change* in investment. To produce goods, a firm needs equipment. Imagine that a machine is capable of producing 1,000 tablet computers per week. Demand for computer tablets is currently 800. A rise in demand for computer tablets of up to 200 is capable of being met without any further investment in new machinery. However, if the rate of growth of demand continues to rise, it may be necessary to invest in a new machine.

Imagine that in year 1, demand for computer tablets rises by 10 per cent to 880. The business can meet this demand through existing equipment. In year 2, demand increases by 20 per cent and is now 1,056. The existing capacity of the machine means that this demand cannot be met but the shortage is only 56 units so the firm decides that it might increase price rather than invest in a new machine. In year 3, demand rises by a further 25 per cent. Demand is now 1,320 but the machine is only capable of producing a maximum

To produce goods, a firm needs equipment which can meet customer demands. For example, the rate of growth of demand for tablet computers continues to rise, firms producing them may find it necessary to invest in a new machine – this is called the ‘accelerator principle’.



of 1,000 tablet computers. The firm decides to invest in a new machine. The manufacturers of the new machine will therefore see a rise in their order books as a result of the increase in demand. An increase in demand of 25 per cent has led to an ‘accelerated’ rise in investment of 100 per cent. Investment is a component of aggregate demand and so economists are interested in the way investment adjusts to changes in demand in the economy. As this brief example shows, the relationship between an increase in demand and an increase in investment is not a simple one.

A Formula for the Spending Multiplier

A little algebra permits us to derive a formula for the size of the multiplier effect that arises from consumer spending. An important number in this formula is the **marginal propensity to consume (MPC)** – the fraction of extra income that a household consumes rather than saves. For example, suppose that the marginal propensity to consume is $\frac{3}{4}$. This means that for every extra pound or euro that a household earns, the household spends $\frac{3}{4}$ of it and saves $\frac{1}{4}$. The **marginal propensity to save (MPS)** is the fraction of extra income that a household saves rather than consumes. With an MPC of $\frac{3}{4}$, when the workers and owners of Nucelec earn €10 billion from the government contract, they increase their consumer spending

marginal propensity to consume (MPC) the fraction of extra income that a household consumes rather than saves
marginal propensity to save the fraction of extra income that a household saves rather than consumes

by $\frac{3}{4} \times \text{€}10$ billion, or $\text{€}7.5$ billion. (You should see from the above that the $MPC + MPS = 1$. The formula below can also be expressed in terms of the MPS as a result.)

To gauge the impact on spending of a change in government purchases, we follow the effects step-by-step. The process begins when the government spends $\text{€}10$ billion, which implies that national income (earnings and profits) also rises by this amount. This increase in income in turn raises consumer spending by $MPC \times \text{€}10$ billion, which in turn raises the income for the workers and owners of the firms that produce the consumption goods. This second increase in income again raises consumer spending, this time by $MPC \times (MPC \times \text{€}10 \text{ billion})$. These feedback effects go on and on.

To find the total impact on the demand for goods and services, we add up all these effects:

| | |
|--------------------------------|---|
| Change in government purchases | = $\text{€}10$ billion |
| First change in consumption | = $MPC \times \text{€}10$ billion |
| Second change in consumption | = $(MPC)^2 \times \text{€}10$ billion |
| Third change in consumption | = $(MPC)^3 \times \text{€}10$ billion |
| • | • |
| • | • |
| • | • |
| Total change in demand | = $(1 + MPC + MPC^2 + MPC^3 + \dots) \times \text{€}10$ billion |

Below, '...' represents a pattern of similar terms. Thus, we can write the multiplier as follows:

$$\text{Multiplier} = (1 + MPC + MPC^2 + MPC^3 + \dots)$$

This multiplier tells us the demand for goods and services that each euro of government purchases generates.

To simplify this equation for the multiplier, recall from your school algebra that this expression is an infinite geometric series. For x between -1 and 1 :

$$1 + x + x^2 + x^3 + \dots$$

The sum of this series as the number of terms tends to infinity is given by:

$$\frac{1}{1 - x}$$

In our case, $x = MPC$. Thus:

$$\text{Multiplier} = \frac{1}{(1 - MPC)}$$

We have said that the $MPC + MPS = 1$ so the multiplier can also be expressed as:

$$\text{Multiplier} = \frac{1}{MPS}$$

For example, if the MPC is $\frac{3}{4}$, the multiplier is $\frac{1}{(1 - \frac{3}{4})}$, which is 4. In this case, the $\text{€}10$ billion of government spending generates $\text{€}40$ billion of demand for goods and services.

This formula for the multiplier shows an important conclusion: the size of the multiplier depends on the marginal propensity to consume. While an MPC of $\frac{3}{4}$ leads to a multiplier of 4, an MPC of $\frac{1}{2}$ leads to a multiplier of only 2. Thus, a larger MPC means a larger multiplier. To see why this is true, remember that the multiplier arises because higher income induces greater spending on consumption. The larger the MPC is, the greater is this induced effect on consumption, and the larger is the multiplier.

The MPC determines the slope of the consumption element of the planned expenditure function.

Other Applications of the Multiplier Effect

Because of the multiplier effect, a euro of government purchases can generate more than a euro of aggregate demand. The logic of the multiplier effect, however, is not restricted to changes in government

purchases. Instead, it applies to any event that alters spending on any component of planned expenditure – consumption, investment, government purchases or net exports.

For example, suppose that a recession overseas reduces the demand for Ireland's net exports by €1 billion. This reduced spending on Irish goods and services depresses Ireland's national income, which reduces spending by Irish consumers. If the marginal propensity to consume is $\frac{3}{4}$ and the multiplier is 4, then the €1 billion fall in net exports means a €4 billion contraction in output.

As another example, suppose that a stock market boom increases households' wealth and stimulates their spending on goods and services by €2 billion. This extra consumer spending increases national income, which in turn generates even more consumer spending. If the marginal propensity to consume is $\frac{3}{4}$ and the multiplier is 4, then the initial impulse of €2 billion in consumer spending translates into an €8 billion increase in aggregate demand.

The multiplier is an important concept in macroeconomics because it shows how the economy can amplify the impact of changes in spending. A small initial change in consumption, investment, government purchases or net exports can end up having a large effect on aggregate demand and, therefore, on the economy's production of goods and services.

Autonomous spending is also an important concept in this analysis. The amount spent in each successive 'round' of spending is termed *induced expenditure*. The multiplier showed how the eventual change in income would be determined by the size of the *MPC* and the *MPS*. The higher the *MPC* the greater the multiplier effect.

However, in an open economy with government, any extra €1 is not simply either spent or saved, some of the extra income may be spent on imported goods and services or go to the government in taxation – withdrawals from the circular flow. Withdrawals (*W*) from the circular flow are classed as endogenous as they are directly related to changes in income. Withdrawals are saving (*S*), taxation (*T*) and imports (*M*).

We also have to take into consideration injections to the circular flow of income. Governments receive tax revenue (a withdrawal from the circular flow) but use it to spend on the goods and services they provide for citizens (an injection into the circular flow); firms earn revenue from selling goods abroad (exports) which are an injection into the circular flow and firms, as we have seen in Chapter 23, use savings (a withdrawal) as a source of funds to borrow for investment (an injection). Injections into the circular flow are exogenous – they are not related to the level of output or income – and are investment (*I*), government spending (*G*) and export earnings (*X*).

The slope of the expenditure line as a whole, therefore, will be dependent on how much of each extra €1 is withdrawn. For each additional €1 of income, some will exit the circular flow of income in taxation, some through savings and some through spending on imports. The marginal propensity to taxation (*MPT*) is the proportion of each additional €1 of income taken in taxation by the government and the marginal propensity to import (*MPM*) the proportion of each additional €1 of income spent on goods from abroad. When we take into consideration the fact that each extra €1 in income is not disposable income, i.e. not all available for consumption, the multiplier effect when considering the *marginal propensity to withdraw (MPW)* will be much lower than if we were simply considering the *MPC* alone in any increase in income.

We can restate the formula for the multiplier (*k*) in an open economy with a government as:

$$k = \frac{1}{MPS + MPT + MPM}$$

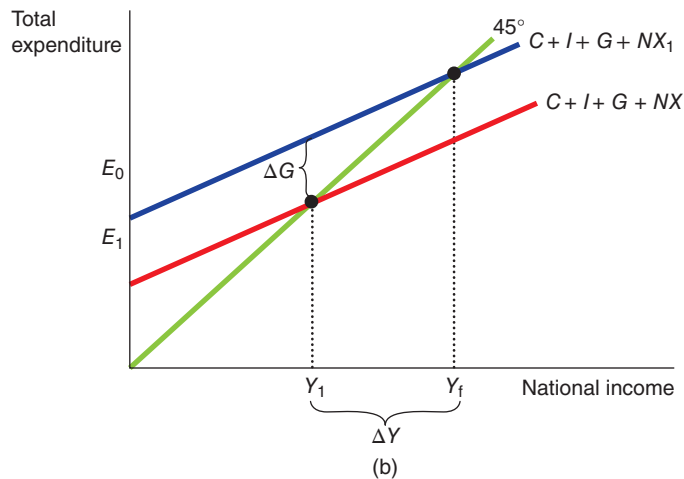
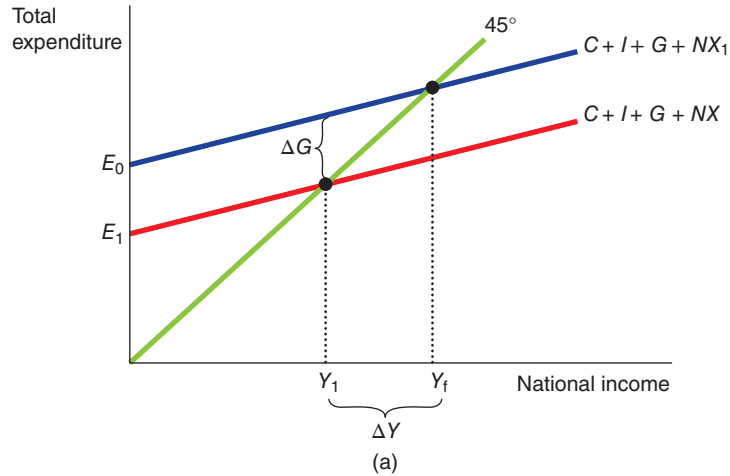
Or:

$$k = \frac{1}{MPW}$$

The size of the *MPW* will determine the slope of the expenditure line: the steeper the slope of the expenditure line the greater the size of the multiplier, as shown in Figure 31.2.

FIGURE 31.2**The Slope of the Expenditure Line and Changes in Autonomous Expenditure**

Panel (a) shows a relatively shallow expenditure line which would mean that the marginal propensity to withdraw would be high and the value of the multiplier was relatively low. The impact on national income (ΔY) of a change in government spending (ΔG) would be more limited in comparison to the effect as shown in panel (b) where the expenditure line is much steeper reflecting a higher value of the multiplier where the MPW was relatively low. In this case it takes a smaller rise in government spending to achieve the same increase in national income.

**Equilibrium of Planned Withdrawals and Injections**

Seeing the economy from the perspective of withdrawals and injections is helpful to understand how demand management might work. Let us start with the national income identity:

$$\text{Output} \equiv \text{Expenditure} \equiv \text{Income}$$

We know from Chapter 23 that in a closed economy $S = I$ with savings a withdrawal and investment an injection. We can also extend this analysis to state that in equilibrium, in an open economy with a government, that planned withdrawals would equal planned injections:

$$\text{Planned } S + T + M = \text{Planned } I + G + X$$

At this point all the output being produced by the economy would be 'bought' by households and firms. However, if actual withdrawals are greater than planned injections then the economy would be experiencing a deficiency in demand. For example, assume that full employment output, (Y), is €120 billion. Withdrawals are a function of income; assume that S , T and M have the following values:

$$\begin{aligned} S &= 0.1Y \\ T &= 0.2Y \\ M &= 0.2Y \end{aligned}$$

Such that if income increased by €1, savings would change by $0.1 \times 1 = 10$ cents and so on.

Now assume that investment was €20 billion, government spending also €20 billion and the value of exports was €10 billion.

Given these figures the equilibrium level of national income would be:

$$\begin{aligned} \text{Planned } S + T + M &= \text{Planned } I + G + X \\ 0.1Y + 0.2Y + 0.2Y &= 20 + 20 + 10 \\ 0.5Y &= 50 \\ Y &= 100 \end{aligned}$$

This equilibrium is below the level of full employment output by €20 billion. The government could manage demand to achieve full employment output in different ways. It could increase its spending by €10 billion and through the multiplier effect see Y rise to €120 billion.

$$\begin{aligned} \text{Planned } S + T + M &= \text{Planned } I + G + X \\ 0.1Y + 0.2Y + 0.2Y &= 20 + 30 + 10 \\ 0.5Y &= 60 \\ Y &= 120 \end{aligned}$$

In an open economy, the government might not simply target full employment but might wish to reduce net exports. For example, assume $Y = €120$ billion, the value of exports is given as €10 billion but the value of imports would be $0.2Y$ and so would be €24 billion (0.2×120). If the government wanted to reduce net exports to zero it might cut government spending but change tax rates so that the marginal propensity to tax increased. If government spending was cut to 5 and the marginal propensity to tax raised to 0.4 then the government could achieve zero net exports:

$$\begin{aligned} \text{Planned } S + T + M &= \text{Planned } I + G + X \\ 0.1Y + 0.4Y + 0.2Y &= 20 + 5 + 10 \\ 0.7Y &= 35 \\ Y &= 50 \end{aligned}$$

It is clear that the policy to reduce net exports has had a severe effect on national income. One of the features of demand management is that it is possible for governments to use fiscal policy to achieve desired output levels but there will be consequences for other areas of the economy which may have more long-term effects. If government sets policy to achieve a reduction in unemployment through boosting the economy, then net exports would fall and the country would be running a trade deficit ($NX < 0$). We saw in Chapter 29 that trade influences net capital outflow and the exchange rate. Changes in the exchange rate affect the competitiveness of firms in the economy and so even though the government might reduce unemployment, its policy causes longer term effects which might cause it to have to alter policy to deal with these effects (such as a currency crisis) which in turn might reduce income and increase unemployment again.

Demand management seemed to work in many countries in the period following World War II up to the mid-1960s but from that point other problems emerged which cast doubt on the viability of demand management which led to Keynesian policies falling out of favour.

THE IS AND LM CURVES

The Keynesian cross gives us a picture of the economy in short-run equilibrium. (Note, if you access a copy of Keynes' *General Theory* you might be surprised to see a complete absence of Keynesian cross diagrams. The use of these diagrams to explain Keynesian ideas was developed by later economists to help portray Keynes' ideas.) In equilibrium, planned expenditure equals income, ($E = Y$). This equilibrium is referred to as equilibrium in the *goods market*. We have also seen, in Chapter 27, how equilibrium in the money market is given by the intersection of the demand for money and the supply of money. We need to consider here the concept of **real money balances** – what money can actually buy or the real value of money given the ratio of the money supply (M) to the price level $P \left(\frac{M}{P} \right)$

real money balances what money can actually buy given the ratio of the money supply to the price level M/P

The goods market and the money market are both interrelated with the linking factor being the interest rate. Following Keynes' analysis of the goods market and the money market (via the liquidity preference theory which we will look at in more detail in a later chapter), Nobel Prize winning economist John Hicks developed a theory that described the links between the two and showed how changes in both fiscal and monetary policy could be analysed. The framework for this analysis is known as the IS-LM model.

IS-LM describes equilibrium in these two markets and together determines a *general equilibrium* in the economy. General equilibrium in the economy occurs at the point where the goods market and money market are both in equilibrium at a particular interest rate and level of income. The remainder of this chapter will provide an introduction to the IS-LM model. The model forms the basis of many intermediate courses in macroeconomics although some have argued that it is now outdated and fails to represent how the modern economy works, particularly since the financial crisis. We will look at an alternative representation of the model that seeks to take into account some of these objections.

Regardless of the debates about the validity of the model, it does represent a useful way of understanding how the goods and money markets interact and as an exercise in analytical thinking is helpful in seeing the effects of monetary and fiscal policy on the macroeconomy.

IS stands for investment and saving; LM stands for liquidity and money. The thing linking these two markets is the rate of interest (i).

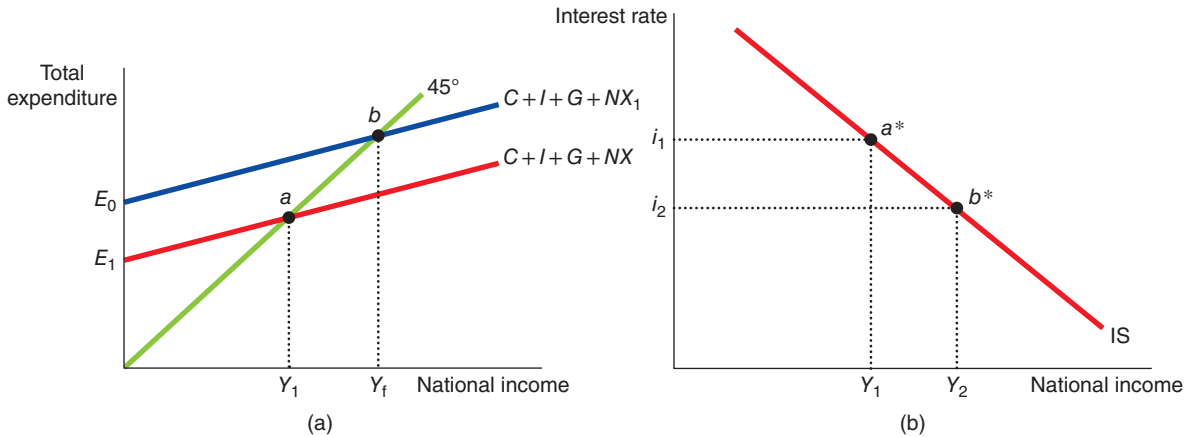
The IS Curve

The IS curve shows the relationship between the interest rate and level of income (Y) in the goods market. In Figure 31.3, panel (a) shows the Keynesian cross diagram from Figure 31.2, with equilibrium point a where the expenditure line $C + I + G + NX$ crosses the 45° line. Panel (b) shows the IS curve. On the vertical axis is the rate of interest and on the horizontal axis is output (national income). The equilibrium point a in panel (a) is associated with a rate of interest i_1 . This is plotted as point a^* on panel (b). If interest rates fall then the expenditure line shifts upwards to the left and there will be a new equilibrium point b where the expenditure line $C + I + G + NX_1$ crosses the 45° line. This is plotted as b^* on panel (b) showing the equilibrium of the goods market at a lower interest rate associated with a higher level of national income. If we connect these two points we get the IS curve. The curve connects all possible points of equilibrium in the goods market associated with a particular interest rate and level of national income.

The IS curve shows an inverse relationship between the interest rate and output – a fall in interest rates leads to a rise in income and vice versa. The rise in income will be dependent on the size of the interest rate change and the size of the multiplier. The slope of the IS curve is determined by the responsiveness of consumption and investment ($C + I$) to changes in interest rates. This is important because it leads to different outcomes; where economists tend to disagree is the *extent* to which $C + I$ are responsive to changes in interest rates rather than any disagreement about the fundamental relationship. The more responsive $C + I$ are, the flatter the IS curve.

FIGURE 31.3**The IS Curve**

The IS curve is derived from the Keynesian cross diagram and shows all possible points of equilibrium in the goods market associated with a particular interest rate and level of income. In panel (a) initial equilibrium is where the $C + I + G + NX$ line crosses the 45° line at point a . This point is plotted on the IS curve as point a^* . An increase in $C + I + G + NX$ to $C + I + G + NX_1$ shows a new equilibrium point in the goods market, b , which is plotted on the IS curve as b^* . These two points are connected to form the IS curve.



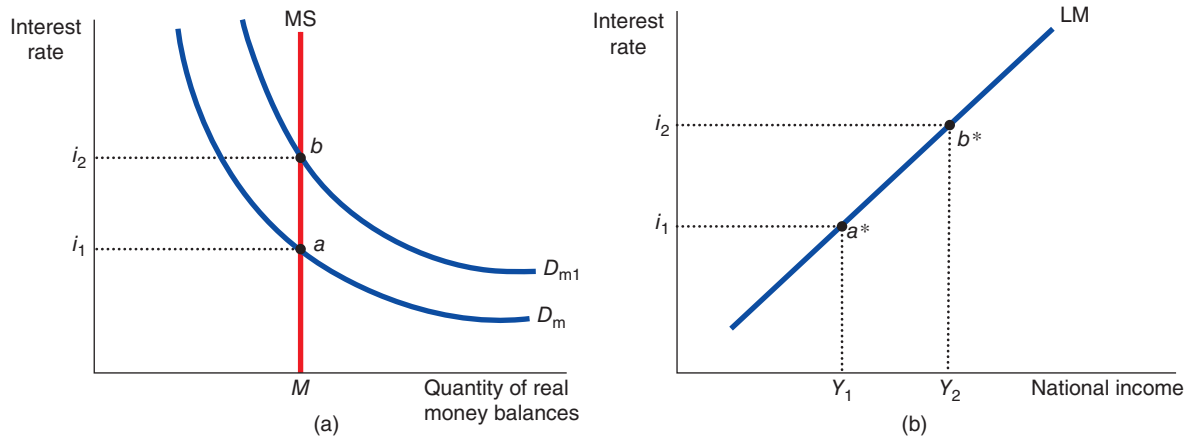
Shifts in the IS Curve Shifts in the IS curve come about as a result of changes in autonomous expenditure. If, for example, government spending rises this occurs independent of any change in interest rates. A rise in autonomous spending would be associated with a shift in the IS curve to the right – the prevailing interest rate would now be associated with a higher level of income. Equally, if autonomous spending fell then the IS curve would shift to the left showing a lower level of income at the prevailing interest rate.

The LM Curve

The LM curve shows all points where the money market is in equilibrium given a combination of the rate of interest and national income. In Figure 31.4, panel (a) shows the money market with the demand for money inversely related to the interest rate. The money supply is shown as a vertical line and it is assumed that the money supply is fixed by the central bank. Equilibrium in the money market is where the demand for money D_m intersects the money supply curve, M_s , at point a in panel (a) at interest rate i_1 and a quantity of real money balances (M). Panel (b) shows the LM curve with the interest rate on the vertical axis and national income on the horizontal axis. The equilibrium point a in the money market is plotted as point a^* in panel (b). Increases in income will have an effect on the demand for money and assuming the money supply is fixed, will affect the equilibrium interest rate. Assume that national income rises; the demand for money curve in panel (a) would shift to the right to D_{m1} indicating that the public wish to hold higher money balances at all interest rates. At the prevailing interest rate the demand for money is now higher than the supply of money and so the interest rate would rise. The new equilibrium in the money market is given as point b and this is plotted on the LM diagram as point b^* . If we connect the two points we get the LM curve. The LM curve plots all combinations of interest rates and national income where the money market is in equilibrium.

FIGURE 31.4**The LM Curve**

The LM curve shows all points where the money market is in equilibrium given a combination of the rate of interest and national income. In panel (a), the money market is in equilibrium where the demand for money (D_m) equals the supply of money (MS) at point a. This point is plotted on the LM curve in panel (b) as point a^* . An increase in the demand for money causes a shift of the curve to the right to D_{m1} with a new equilibrium point of b. This is plotted on panel (b) as point b^* and the points connected to form the LM curve.



The LM curve has a positive slope showing that an increase in income is associated with an increase in the interest rate and vice versa. The slope of the LM curve will be dependent on the responsiveness of the demand for money to changes in interest rates. Again, the extent of this relationship is often a point of disagreement among economists.

Shifts in the LM Curve The LM curve can shift if the central bank expands or contracts the money supply (we will say more about how this might happen later in the chapter). Assuming income is held constant, a rise in the money supply (for example) will cause interest rates to fall and a new equilibrium will be reached at a given level of income. This would be associated with a shift of the LM curve downwards to the right showing a new combination of income and interest rate at which the money market is in equilibrium.

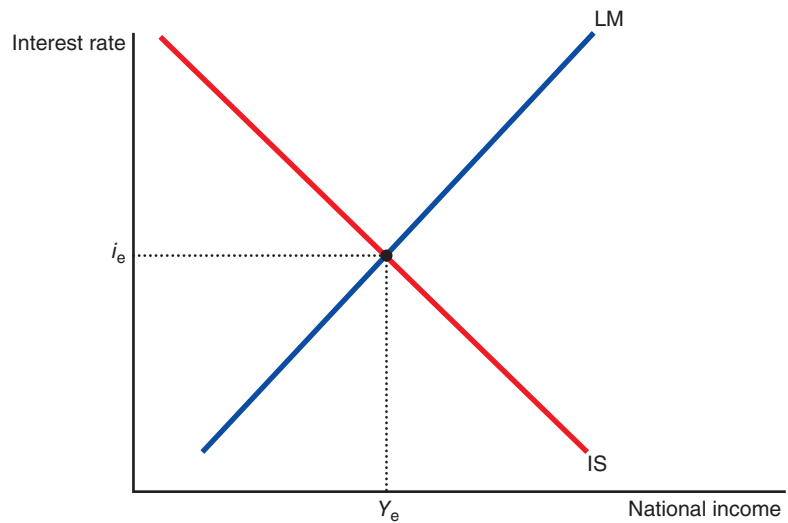
GENERAL EQUILIBRIUM USING THE IS-LM MODEL

Equilibrium is found where the IS curve intersects the LM curve. Remember that any point on either curve describes a point of equilibrium in the goods market and the money market at a rate of interest and level of national income. In Figure 31.5, the point where the IS curve intersects the LM curve gives a point where both markets are in equilibrium at an interest rate i_e and a level of national income Y_e . Hence, it follows that at this point planned expenditure equals actual expenditure ($E = Y$), and the demand for money equals the supply of money ($D_m = S_m$).

Having established this general equilibrium, we can use the model to analyse the impact of fiscal and monetary policy changes in an attempt to stabilize the economy and how these two policies are interrelated. Further analysis of both policies will be covered in the next chapter but this uses our model of IS-LM. The detail of IS-LM analysis is beyond the scope of this book, being found in most intermediate courses in macroeconomics. However, the remainder of this chapter will introduce some of the key implications of the model.

FIGURE 31.5**General Equilibrium**

Equilibrium in the economy is found where the IS curve intersects with the LM curve. At this point both the goods market and the money market are in equilibrium at a particular interest rate (i_e) and level of national income (Y_e).

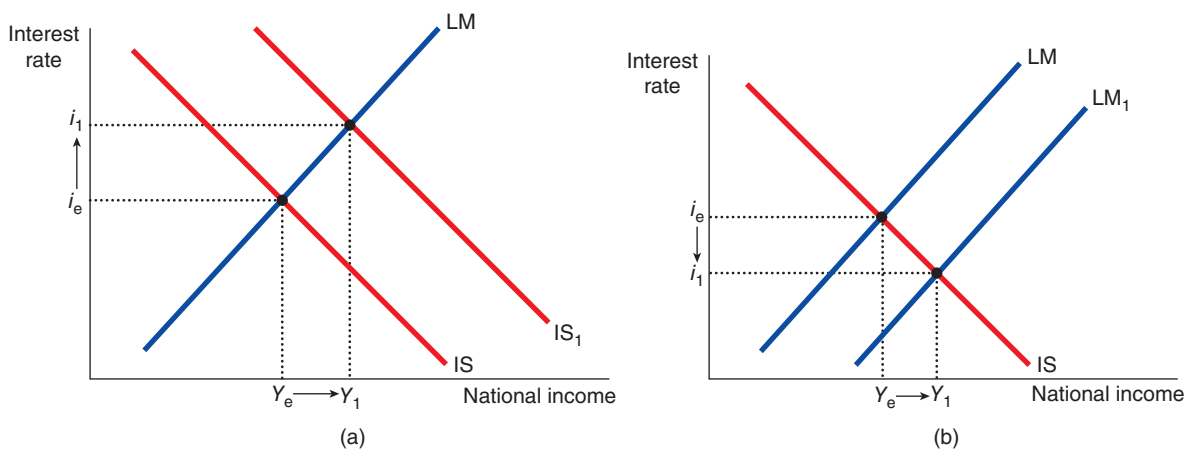


The Effect of a Change in Fiscal Policy

Assume that the government chooses to increase spending to boost economic activity. This increase in autonomous expenditure shifts the IS curve to the right as shown in panel (a) of Figure 31.6. The result is that national income will rise but there will also be an increase in interest rates. A similar outcome would occur if the government chose to cut taxes as the means of boosting the economy. The result of either policy would be dependent on the marginal propensity to withdraw and the size of the multiplier. The opposite would occur if the government chose to cut spending or increase taxes – national income and interest rates would both fall.

FIGURE 31.6**The Effects of Fiscal and Monetary Policy**

In panel (a) a rise in government spending shifts the IS curve to the right resulting in a new equilibrium with a higher interest rate and level of national income. In panel (b) an increase in the money supply would shift the LM curve to the right and a new equilibrium would result in a lower interest rate and higher level of national income.



The Effect of a Change in Monetary Policy

If the central bank decided to expand the money supply the LM curve would shift to the right to LM_1 as shown in panel (b) of Figure 31.6. The new equilibrium would lead to a lower interest rate and a higher level of national income. The reverse outcome would occur if the central bank tightened monetary policy by reducing the money supply.

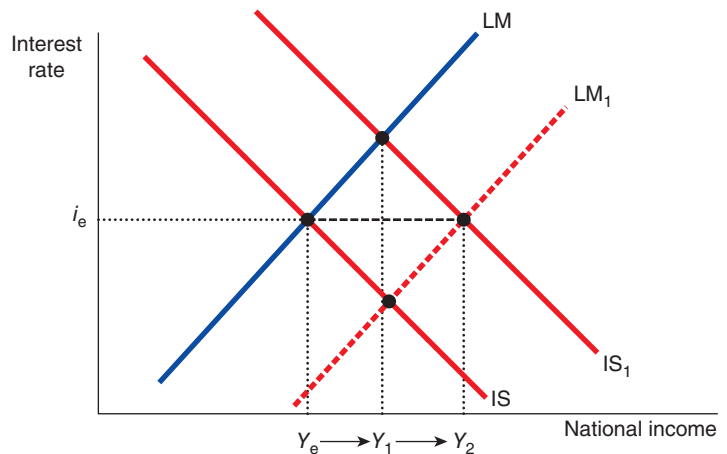
Fiscal and Monetary Policy Interactions In reality, central banks do not act totally in isolation of government even if they are independent. Central banks will be aware of what governments are doing as we will see when we look at the Case for an Active Stabilization Policy in Chapter 33. This presents a model to analyse the response of the central bank. The Bank of England, European Central Bank and the Federal Reserve have a responsibility to maintain price stability and this may be presented in the form of a target for inflation. Central banks will be monitoring the effect of fiscal policy changes on the economy and how these changes might affect inflationary pressures. These inflationary pressures can be influenced by the central bank's control over short-term interest rates through the rate at which it lends to the financial system. Governments may wish to implement fiscal policy with the aim of influencing unemployment, for example. Such a policy may have effects on inflationary pressures which the central bank wants to nullify.

Let us assume that the government reduces taxation to encourage more people to take jobs in the economy or to increase spending through consumers having more disposable income. The IS curve would shift to the right as shown in Figure 31.7 and national income and interest rates would rise. If the central bank wants to keep interest rates constant it must expand the money supply. By doing so the LM curve shifts to the right to LM_1 , national income would rise further than if the central bank had not acted, to Y_2 and the interest rate will remain at its initial level. If the bank had not altered the money supply then the effects of the reduction in tax would have been partially offset by a rise in interest rates which would have curbed spending.

FIGURE 31.7

Maintaining Interest Rates Constant Following a Rise in the IS Curve

A shift in the IS curve to the right would, without central bank action, lead to a rise in the interest rate and in national income. If the central bank wants to maintain the interest rate it must increase the money supply and shift the LM curve to the right. The result would be to maintain the interest rate at i_e but the increase in national income would be greater than if the central bank had not acted.



If government had increased taxes then the IS curve would shift to the left and both national income and interest rates would fall. If the central bank wants to keep interest rates constant it must reduce the money supply and the result would be that the fall in national income would be compounded. If the central bank wanted to avoid this outcome then it could expand the money supply and interest rates would fall. This would help to offset the shift in the IS curve and reduce the impact on national income.

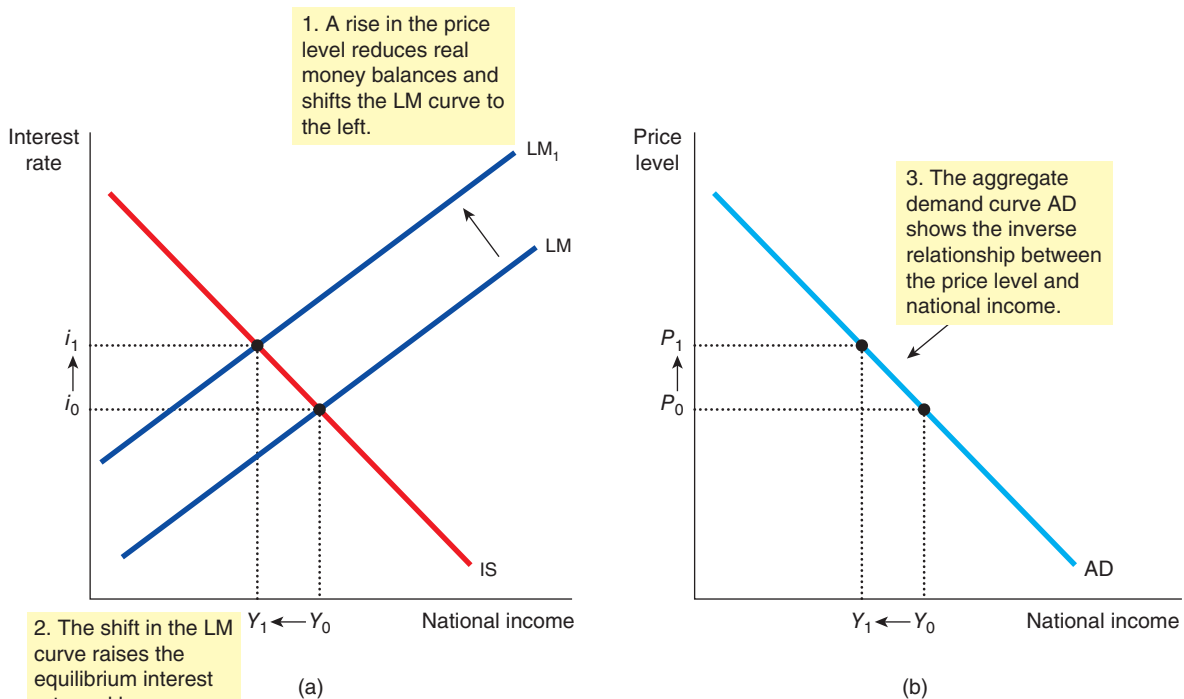
FROM IS-LM TO AGGREGATE DEMAND

It is a short step from using this model to the aggregate demand and aggregate supply model that we will use to analyse changes in the economy later in the book. Remember that the supply of real money balances is what money can actually buy given by M , the money supply divided by P , the price level $\left(\frac{M}{P}\right)$.

Assume the average price of a unit of output in the economy is €10 and the money supply is €100 billion. The supply of real money balances is $\frac{100}{10} = 10$, that is at the current price level the supply of money in the economy can buy 10 units of output. If the average price of a unit of output in the economy rises to €20 and the money supply is constant, real money balances will fall to 5 units ($\frac{100}{20} = 5$). This fall in the supply of real money balances shifts the LM curve to the left as shown in panel (a) of Figure 31.8. The result is that interest rates rise and national income falls. There is, therefore, an inverse relationship between the price level and national income. The aggregate demand curve is derived by plotting the relationship between national income and the price level as shown in panel (b) of Figure 31.8. The aggregate demand curve slopes downwards from left to right because of the inverse relationship between the price level and national income.

FIGURE 31.8**Deriving the Aggregate Demand Curve**

In panel (a) a rise in the price level reduces real money balances and shifts the LM curve to the left to LM_1 . This leads to a rise in the equilibrium interest rate to i_1 and a fall in national income from Y_0 to Y_1 . The inverse relationship between the price level and national income is plotted in panel (b). A rise in the price level from P_1 to P_2 leads to a fall in national income from Y_0 to Y_1 . The aggregate demand curve slopes downwards from left to right.



Shifts in the Aggregate Demand Curve If we assume the price level remains constant, a change in national income in the IS-LM model will result in a shift in the aggregate demand curve. Changes in both fiscal and monetary policy, assuming a given constant price level, will cause the aggregate demand curve to shift. If the government imposes an austerity package which seeks to cut government spending and raise taxes, at a given price level the IS curve will shift to the left and national income will fall. At the given price level, aggregate demand in the economy will now be less and the aggregate demand curve will shift to the left as shown in panel (a) of Figure 31.9.

If the central bank expands the money supply (possibly through a programme of asset purchasing or quantitative easing) the LM curve will shift to the right and national income will rise. At the given price

level, the aggregate demand curve will shift to the right showing a higher level of national income at the given price level as shown in panel (b) of Figure 31.9.

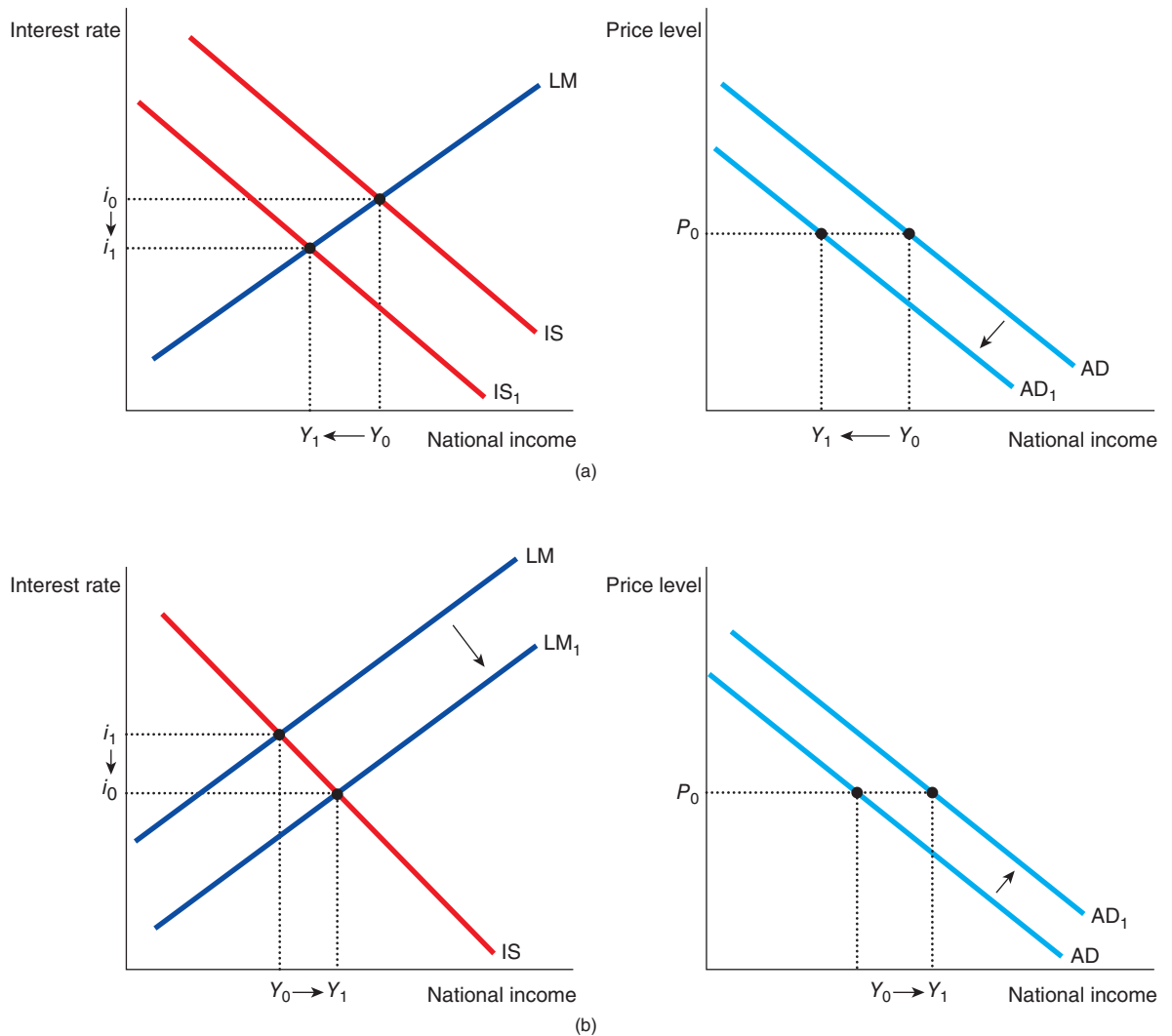
A loosening of fiscal policy (increased government spending and/or lower taxes) and a reduction in the money supply (a tightening of monetary policy) will have the opposite effect to that described in Figure 31.9 respectively.

FIGURE 31.9

Shifts in the Aggregate Demand Curve as a Result of Monetary and Fiscal Policy

Panel (a) represents a situation where the government tightens fiscal policy which shifts the IS curve to the left and reduces national income. At a given price level the aggregate demand curve shifts to the left and a lower level of national income is associated with the given price level.

Panel (b) represents a situation where the central bank loosens monetary policy which causes the LM curve to shift to the right and lowers national income. At the given price level the aggregate demand curve shifts to the right with a higher level of national income.



Criticisms of IS-LM and the Romer Model

This short introduction to IS-LM analysis and the effect of changes in fiscal and monetary policy helps to highlight a number of important issues that you may start to consider as your study of economics

moves to the next level. The effects of changes in fiscal and monetary policy are dependent on a number of factors related to the slope of the IS and LM curves, their relative position and how far each shifts in response to changes in policy. It is entirely possible that a change in fiscal policy can be countered by a change in monetary policy that leaves national income unchanged. The potential outcomes are many and economists try to pinpoint more accurately what changes in policy will mean for the economy as a whole. This means conducting research to quantify such changes. The outcome of such research will be dependent on the value of variables that economists input into their models and, as has already been stated, most economists agree on the fundamentals of the model but disagree on the relative strength of factors affecting the variables in the model.

One example of how the model has developed is the role which microeconomic analysis plays in understanding the macroeconomy. Some economists argue that microeconomic principles cannot be divorced from the macroeconomy. We have seen in Chapter 30 how changing economic conditions may lead to wages and prices adjusting at different rates and slowly as a result. If national income falls, for example, then the assumption might be that prices and wages in the economy would also fall to help bring markets back into equilibrium. As we will see in the next chapter, these sticky wages and prices may lag behind the reduction in economic activity as firms may be forced into trying to maintain cash flow rather than seeking to expand market share. As prices are sticky, sales fall and firms cut back output which further impacts on economic activity. Price stickiness seemed to be counter to the prevailing assumption about rational behaviour.

Professor Mankiw was one of the economists who helped to reconcile the idea of sticky prices and menu costs with rational behaviour in his paper 'Small menu costs and large business cycles: A macroeconomic model of monopoly', published in *The Quarterly Journal of Economics* in 1985. The debate over IS-LM has continued and has generated large amounts of valuable and interesting research which has helped build our understanding of the economy as a whole. It still has its critics, however.

Indeed, one such disagreement focuses on a central assumption of the model itself. Some higher education institutions have questioned the value of teaching IS-LM at all because they argue that the world is now a very different place to the one Hicks knew back in the 1930s when he first developed Keynes' ideas. One of the major criticisms is that central banks no longer control the money supply but instead set interest rates. Setting interest rates is seen as leading to more stable economic conditions than policies targeting the money supply. Attempts to control the money supply have proved to be difficult and so targeting the interest rate is seen as being a more viable option to achieve policy objectives.

In Chapter 26 we looked at how central banks set interest rates through open market operations. If the central bank wishes to reduce interest rates, its traders will be instructed to buy bonds. Banks and financial institutions who sell these bonds will receive funds in return which will effectively expand the money supply. This in turn shifts the LM curve to the right and so interest rates fall. If the central bank increases interest rates the opposite occurs; traders will be instructed to sell bonds and thus take funds out of the banking system, reducing the money supply. The LM curve shifts to the left and interest rates rise.

One of Professor Mankiw's close colleagues, David Romer (indeed they are more than simply colleagues given that each was best man at the other's wedding), has suggested an approach termed the IS-MP model which attempts to build on the IS-LM model to reflect how central banks and the economy work today. The assumption in the model is that central banks adjust the money supply as outlined above to generate the interest rate that they want. The interest rate is adjusted in accordance with the inflation target that the central bank is working with or, as is the case with the Bank of England, has been set by the government. In the IS-LM model the money supply is assumed to be exogenous (determined by factors outside the model). In the IS-MP model, the monetary policy reaction function is exogenous. Romer assumes that when output rises, the central bank increases interest rates to dampen inflationary pressures and reduces interest rates when output falls to maintain the price level at its target. National income is a positive function of the interest rate, therefore. Romer plots this upward sloping relationship as the MP curve. The MP function is assumed to be exogenous but in reality both the money supply and the MP function can and do change in response to economic activity and events.

The MP function, therefore, takes into account the fact that central banks now target inflation and set interest rates to achieve this goal rather than simply assuming government (or a central bank) sets the money supply and that interest rates adjust to balance this supply of money with the demand for money. Crucially, Romer suggests that changes in inflation can cause a shift of the MP curve. If the central bank increases interest rates the money supply will fall which will affect the price level and expectations on

inflation. Equally, if the central bank cuts interest rates the money supply will rise and expectations on inflation will also change as a result.

It is at this point where the microeconomic element of the analysis takes on some importance. Expectations of price changes may not match the reality because of the extent of price stickiness. If we assume that prices are completely sticky (i.e. the price level is fixed) it will not change when the money supply changes. Expectations on inflation will, therefore, be zero. If the money supply rises then the supply of real money balances, $\frac{M}{P}$, also rises and is greater than the demand for real money balances. As the money market is now out of equilibrium we might expect the interest rate to fall but it could also be that the level of income could rise or a combination of the two might occur. There will be a movement along the IS curve and a fall in interest rates will be accompanied by a rise in national income. This implies that the central bank can directly control the real interest rate by adjusting money supply appropriately to achieve the interest rate it desires.

If prices adjusted instantaneously then a change in the money supply would not affect the supply of real money balances because the ratio $\frac{M}{P}$ would hold as before at the rate of interest and level of income; in other words the money market would remain in equilibrium. There would be no movement along the IS curve and the central bank would not be able to affect the real interest rate.

In reality prices are not fixed, but the speed with which prices adjust to changes in economic conditions will vary – some will adjust relatively quickly, others will take much longer and will be sticky. These sticky prices will influence the ratio of the money supply to the price level (the supply of real money balances) and so there may be some expectations of inflation which will exist in the economy because M will rise by a greater proportion than P . An increase in the money supply will raise expected inflation and vice versa. If prices are sticky then an increase in the nominal money supply will cause the money market to move out of equilibrium but expected inflation will affect the real interest rate. The nominal interest rate may have to be higher as a result.

The extent to which prices are sticky, therefore, has an important influence on the way in which a central bank can influence the interest rate to achieve its inflation targets. Such an analysis raises interesting questions when interest rates are reduced as was the case in response to the financial crisis and the global recession in 2008–2009. The UK, Europe and the US saw their central banks reducing interest rates to historically low levels. To reduce interest rates the central bank instructs its traders to buy bonds and the money supply will rise. Increases in the money supply may be associated with an increase in the price level as noted in Principle 9 of the *Ten Principles of Economics*. Indeed, some economists have expressed alarm at the scale of quantitative easing conducted by central banks and have predicted hyperinflation. In times where economic activity is restrained, however, expectations of inflation may remain subdued and the money market may remain out of equilibrium as a result. The reason is that at these low interest rates people are willing to hold a greater amount of real money balances without any change in interest rate or output – after all, the interest rate cannot fall much further. This is the *liquidity trap* which is covered in the next chapter and which may imply that monetary policy can have little effect on stimulating economic growth.

SELF TEST Draw diagrams to show the effect on interest rates and the level of national income of: (a) a decision by the government to raise taxes to cut a public deficit; and (b) a decision by a central bank to increase interest rates.

A Return to Keynesianism?

The financial crisis of 2007–2009 led to a period of global recession. In many countries, economic growth has been slow to return to pre-crisis levels and levels of unemployment have been rising in many countries. The severity of the global recession led to calls for Keynesian style fiscal intervention to stimulate economic growth, particularly given the fact that central banks had largely exhausted traditional methods of monetary policy to stimulate growth. This reignited the debate about the value of Keynesian demand management policies and the news media and blogosphere has been alive with, what has sometimes been, bitter divisions between proponents of expansionary fiscal policy and those who focus more on the importance of governments reducing deficits. Indeed, in one instance, the website Wikipedia locked down

a page to prevent editing following a series of edits and counter-edits between those in support of Nobel Prize-winning economist Paul Krugman, a supporter of Keynesian style intervention, and supporters of the Austrian school who tend to be sceptical of the benefits of government intervention and are more inclined to trust markets to sort out economic problems. Whether the 'editing war' was anything to do with Krugman himself is not clear but after a series of repeated cases where text was deleted and replaced by both sides, Wikipedia closed the page to further editing.

Regardless of the debate, a number of countries introduced fiscal stimulus packages in the wake of the financial crisis. This led to questions about the benefits of such packages and even whether they amount to a stimulus. In late 2008 the European Union (EU) announced a fiscal stimulus package of €200 billion; in the UK the Chancellor of the Exchequer had to admit that government borrowing might reach £175 billion and be 79 per cent of GDP in 2013–2014 (in the latter part of 2012 it stood at 73 per cent of GDP). In India public debt as a component of GDP is around 80 per cent; in the US the package was reported to be \$800 billion. China was reported to be injecting \$586 billion and Japan \$275 billion. Critics were arguing that such packages were not enough to bridge the output gap that had widened in most economies. In the US, for example, the gap between potential and actual GDP was estimated by some economists at around \$2 trillion. Other criticisms of fiscal stimulus packages suggested that they were not really what they seemed to be and if they were then they would not bring the benefits that were claimed for them because of crowding out.

For example, in a study of the EU stimulus package in February 2009, David Saha and Jakob von Weizsäcker said, 'It should be recognized that the likely real impact on aggregate demand in the near future may well be more limited than suggested by the headline figures' (<http://aei.pitt.edu/10549/01/UPDATED-SIZE-OFSTIMULUS-FINAL.pdf> accessed 6 October 2009). In their analysis of the fiscal stimulus in Italy, announced as €80 billion, Saha and Weizsäcker conclude that the stimulus was not a stimulus at all but a fiscal tightening amounting to €0.3 billion. In parts of Europe there were concerns that any major fiscal stimulus package would put pressure on the public debt and affect the stability of the euro around which so much of the future prosperity of the EU lies.

One of the concerns about any fiscal stimulus is the extent to which it creates 'real wealth'. Governments may spend more money but on what? If the money is spent on public works – the building of new schools, hospitals, roads and so on, then surely this would boost aggregate demand? To an extent it would but how is this additional spending to be financed? To raise the money governments will either have to tax their citizens more or increase borrowing. Additional spending on the construction of a new road may put money into the pockets of construction companies and its workers, but to fund this the government has to tax other wealth producers, thus offsetting some of the benefits of the stimulus. If the spending goes on additional benefits for those who become unemployed then critics argue that the government is not contributing to wealth creation; these individuals may be supported in times of hardship and may spend their benefits on food and other goods but they are not actually generating any wealth in return for the benefits they receive. Indeed, Keynes himself acknowledged that there are limitations to stimulus packages. In an article in *The Times* newspaper in the UK in 1937, a year after the *General Theory* was published, he wrote:

But I believe that we are approaching, or have reached, the point where there is not much advantage in applying a further general stimulus at the centre ... It follows that the later stages of recovery require a different technique. To remedy the condition of the distressed areas ad hoc measures are necessary. The Jarrow marchers were, so to speak, theoretically correct ... We are in more need today of a rightly distributed demand than of a greater aggregate demand.

Structural changes in the economy might require more long-term investment if improvements in unemployment prospects and growth are to be achieved.

If governments have to borrow more then crowding out may emerge. In this scenario the government is competing with the private sector for funds. Since the supply of loanable funds is finite, if governments take more of these funds it is argued that there will be less available for the private sector. If it is also assumed that the private sector uses such investment funds more efficiently than the public sector then not only do fiscal stimulus packages crowd out private investment they divert investment funds to less productive uses.

Whilst this view is widely accepted, there are those that suggest that extraordinary times call for extraordinary action. One such proponent of this view is Paul Krugman. Krugman has consistently argued that the depth of the financial crisis and global recession was such that a fiscal stimulus was necessary to get

out of a liquidity trap and far from crowding out, fiscal stimuli would lead to 'crowding in'. A liquidity trap occurs when monetary policy is insufficient to generate the economic stimulus necessary to get out of recession. In the US and UK, for example, interest rates were lowered by the Fed and the Bank of England to near zero, the Bank of Japan's key interest rate stood at 0.1 per cent in August 2009 whilst the ECB had rates at 1 per cent.

Krugman argues that private investment is a function of the state of the economy and, given the depth of the global recession, investment has plummeted. Investment, therefore, in this situation is more responsive to product demand than to the rate of interest. If governments applied appropriate fiscal stimuli then this would improve the state of the economy and thus encourage private sector investment. As private sector investment increases then this improves the productive potential and helps bring economies out of recession. Far from crowding out, Krugman argues, the fiscal stimulus would lead to crowding in.

Critics of this view suggest that increased government spending will crowd out private investment to an extent and so any fiscal stimulus has to take into account the loss of the benefits of that private investment in assessing the success of such a policy. The success of a stimulus package in putting economies on a sounder footing and bringing benefits to future generations would be highly dependent on the type of spending carried out by governments worldwide. Spending on new schools and transport infrastructure may improve the future productive potential of the economy and tackle the structural problems that exist in some economies as referred to above. However, the issues of rent seeking and log rolling have to be taken into consideration. Remember that rent seeking occurs where decisions are made leading to resource allocation that maximizes the benefit to the decision-maker at the expense of another party or parties, and log rolling is where decisions may be made on resource allocation to projects that have less importance, in return for the support of the interested party in other decision-making areas. In both cases, it is argued that resource allocation is not as efficient as that carried out by the private sector and this has to be taken into consideration in assessing the benefits of public sector spending as a result of any stimulus package.

CONCLUSION

There is still much debate on whether Keynesian style interventions in the economy to manage demand represent an appropriate and viable response to the economic conditions currently facing many countries in the aftermath of the financial crisis. There is also some debate on the value and relevance of the IS-LM model. What is not in much doubt is that an understanding of the relationship between the goods market and the money market is a useful way of developing a broader understanding of analysing the economy as a whole. This short introduction to such analysis provides some pointers to the IS-LM model and to some of the issues that economists are debating. Depending on the university that you attend, a greater or lesser emphasis may be placed on the IS-LM model. Having some awareness of the model does help to develop a focus on the important connections between the money supply, interest rates and economic activity.

IN THE NEWS



Austerity and Growth

The aftermath of the financial crisis has presented economists with new challenges as many countries try and face problems of high levels of borrowing and debt which they wish to reduce, and faltering economies with weak growth, high unemployment and, in some cases, inflation, stubbornly persisting at levels above target.

Expansionary Austerity or Keynesian Intervention?

One of the legacies of the financial crisis was the wider public realization that governments around the world had been borrowing heavily to finance spending and this spending had helped to boost strong growth throughout much of the latter part of the 1990s and into the noughties. When the crisis hit, consumption spending, a key driver of growth, fell and the recession that followed was deep and long lasting.

Central banks responded to the crisis by cutting interest rates and in many countries historically low interest rates have persisted to the extent that it is not really possible for central banks to try and boost economic growth by traditional monetary policy tools any further. For governments, it is clear that debt levels have to be reduced if the risk of default and the ability of governments to borrow in the future is to be maintained. What is the way to reduce borrowing and debt? Cut government spending and increase taxes is the obvious policy choice and some finance ministers and political leaders are arguing that there is simply no alternative to such fiscal contraction.

Some economists have cited the experience in Ireland as justification for austerity. Ireland had to implement savage fiscal cuts to get financial assistance from the European Union and International

Monetary Fund (IMF). The economy shrank alarmingly but forecasts put growth at 1.6 per cent in 2013 and 2.8 per cent in 2014. The head of the IMF, Christine Lagarde, was keen to praise the Irish government – and the people – for the progress the country has made in dealing with the deficit and the banking system. In a meeting with Irish government ministers in March 2013, Ms Lagarde said:

‘What has been done is huge by any standard. The amount by which the deficit has been reduced, the determination shown in actually implementing a programme has been extraordinary ... It doesn’t mean to say that all has been done but I would say that more than two-thirds of the work has been done in terms of fiscal policies and there is still more work to do on the banking sector to make sure it is safe, solid.’

Such praise might be welcome for some in Ireland but the cost of austerity in terms of the sacrifices made by the people has been considerable – in March 2013, unemployment was over 14 per cent and was not forecast to fall much below 13.5 per cent by the end of 2014.

The experience of Ireland was not a surprise. Keynes argued in the 1930s that such policies would

lead to depression. The economic conditions in many other countries across Europe look rather like those expected in a depression. When government spending is cut, public sector employment falls, government spending on benefits rises and tax revenue falls. The private sector might be able to expand to pick up the unemployed from the public sector (an approach which has been referred to as ‘expansionary austerity’) but in such gloomy economic times, many large corporates are hoarding cash and postponing investment. The size of public sector cuts and the resulting unemployment plus the gloomy news which dampens confidence and leads to negative expectations causes a further slowdown in spending – the cycle reinforces itself. With consumption spending falling, tax revenues from income and corporate taxes fall even further and governments are forced to borrow so the deficit and borrowing does not fall as far as finance ministers forecast. The markets get nervous and push up borrowing rates for sovereign debt which makes it more likely they will default on existing debt. In 1937, Keynes observed: ‘The boom, not the slump, is the right time for austerity at the Treasury.’

What is needed, say some, is actually more borrowing not less. Policies are needed to get consumers spending again and firms investing, and governments can

trigger this process by spending more on capital investment on key infrastructure projects which will improve the future productive capacity of the economy and address structural problems. Some economists argue that it is time to embrace once again the wisdom shown by Keynes and that the severity of the global downturn calls for coordinated global efforts to boost consumer spending and investment. Those economists who have called for a return to Keynesian intervention have used the data on countries like Ireland to point to the failures of austerity.

Robert Skidelsky, Emeritus Professor of political economy at Warwick University in the UK, a biographer of Keynes, noted the difference in the more expansionary approach taken by the Obama administration in the US up to 2013 and the austerity approach adopted by the UK government:

Since May 2010, when US and British fiscal policy diverged, the US economy has grown – albeit slowly. The British economy is currently contracting. Unemployment in the United States has gone down by 1.4 percentage points; in Britain, it has

Robert Skidelsky, Emeritus Professor of political economy at Warwick University in the UK and famed biographer of Keynes, pointed out differences in the more expansionary approach taken by the Obama administration in the US up to 2013 and the austerity approach adopted by the UK government.



gone up by 0.2 percentage points. And despite keeping up stimulus measures, the Obama administration has been more successful in reducing the government deficit – by 2.5 percentage points compared with Osborne’s 1.9 percentage points. (<http://www.newrepublic.com/article/politics/magazine/104220/austerity-keynes-uk-us-skidelsky>).

Skidelsky’s conclusion is that the poor performance of the UK economy and of those in other countries where austerity has been the dominant factor, it is not surprising to those who believe Keynesian

policies have a part to play in dealing with the economic problems of the post-crisis world. What he and other Keynesians are noting, however, is that whilst opinion is gradually changing against expansionary austerity, it may be too late.

Questions

- 1 How would central banks have brought about the considerable cuts in interest rates in the immediate aftermath of the financial crisis? Show the effect of such a policy on aggregate demand using the IS-LM model.
- 2 Use the concept of the multiplier to explain how cuts in government spending and higher taxes are likely to lead to a fall in national income.
- 3 Do you think that the benefits of implementing austerity measures in Ireland are justified? Explain your reasoning.
- 4 Outline why governments such as that in the UK, believe that there is no viable alternative to austerity if longer term sustainable growth is to be achieved.
- 5 To what extent do you agree with the view that ‘expansionary austerity’ has not worked and that the time has come for Keynesian fiscal expansion?

SUMMARY

- Keynes developed *The General Theory* as a response to the mass unemployment which existed in the 1930s.
- He advocated governments intervene to boost demand through influencing aggregate demand.
- The Keynesian cross diagram shows how the economy can be in equilibrium when $E = Y$.
- This equilibrium may not be sufficient to deliver full employment output and so the government can attempt to boost demand to help achieve full employment.
- John Hicks developed Keynes’ ideas in the form of the IS-LM model which shows general equilibrium in the economy.
- The IS (investment–saving) curve shows all points of equilibrium in the goods market at a particular interest rate and level of national income.
- The LM (liquidity–money supply) curve shows points where the money market is in equilibrium at particular rates of interest and level of national income.

- General equilibrium occurs where the IS curve intersects the LM curve. At this interest rate and level of national income both the goods market and money market are in equilibrium.
- Fiscal policy and monetary policy can cause shifts in the IS and LM curves bringing about new equilibrium positions. The outcome will depend on a variety of factors including the response of consumption and investment

to changes in interest rates and the public's response to holding monetary balances as a result of a change in interest rates.

- There have been criticisms that the IS-LM model does not represent the way monetary policy is conducted in modern economies.
- Economists have developed new models to incorporate the changes in policy.

QUESTIONS FOR REVIEW

- 1 Distinguish between planned expenditure and actual expenditure.
- 2 Draw a Keynesian cross diagram to show the effects of a rise in autonomous expenditure on an economy operating below full employment output.
- 3 What is meant by the terms: *inflationary gap* and *deflationary gap*?
- 4 What is the marginal propensity to consume?
- 5 Why does the $MPC + MPS = 1$?
- 6 What is the multiplier? Can the multiplier be negative as well as positive? Explain.
- 7 Explain how the marginal propensity to withdraw affects the outcome of a rise in autonomous expenditure.
- 8 Use diagrams to describe how the IS and LM curves are derived.
- 9 Using the IS-LM model, explain the effect on the economy of a reduction in autonomous expenditure resulting from a cut in public spending by the government.
- 10 A central bank wishes to reduce inflationary expectations by increasing interest rates. Use the IS-MP model to analyse the effect on the economy of such a move.

PROBLEMS AND APPLICATIONS

- 1 What, according to Keynes, was the main reason why recessions and depressions occurred? As a result of identifying this key reason, what did Keynes suggest was an appropriate policy response?
- 2 Explain, using an appropriate diagram, how a deflationary gap can occur and how this gap can be eliminated.
- 3 Suppose economists observe that an increase in government spending of €10 billion raises the total demand for goods and services by €30 billion.
 - a. If these economists ignore the possibility of crowding out, what would they estimate the marginal propensity to consume (MPC) to be?
 - b. Now suppose the economists allow for crowding out. Would their new estimate of the MPC be larger or smaller than their initial one? Explain your answer.
- 4 Suppose the government reduces taxes by €2 billion, that there is no crowding out and that the marginal propensity to consume is 0.75.
 - a. What is the initial effect of the tax reduction on aggregate demand?
 - b. What additional effects follow this initial effect? What is the total effect of the tax cut on aggregate demand?
 - c. How does the total effect of this €2 billion tax cut compare to the total effect of a €2 billion increase in government purchases? Why?
- 5 Assume the economy is in equilibrium. Analyse the effect of a cut in autonomous expenditure on economic activity and the level of unemployment. You should use a diagram to help illustrate your answer.
- 6 What does the IS curve show? What does the LM curve show?
- 7 What determines the slope of the IS curve? What determines the slope of the LM curve? In relation to your answer to these questions, explain why these determinants can be a source of disagreement amongst economists?
- 8 Use IS-LM analysis to explain the following:
 - a. The government institutes significant cuts in public expenditure.
 - b. The central bank institutes an asset purchasing facility which expands the money supply by €300 billion.
 - c. The central bank fears that inflationary pressures are rising and increases interest rates.
 - d. The government increases taxation to try and reduce a large budget deficit.
- 9 Assume that a period of deflation leads to a rise in the supply of real money balances. Explain the effect of this change on the economy using the IS-LM model and then what effect it would have on aggregate demand and why.
- 10 Do you think that Keynes' ideas still have some relevance today? Explain.

32

AGGREGATE DEMAND AND AGGREGATE SUPPLY

We have seen how economic activity can fluctuate over different time periods. What causes short-run fluctuations in economic activity? What, if anything, can public policy do to prevent periods of falling incomes and rising unemployment? When recessions and depressions occur, how can policymakers reduce their length and severity? These are the questions that we take up now.

The variables that we study are largely those we have already seen in previous chapters. They include GDP, unemployment and the price level. Also familiar are the policy instruments of government spending, taxes, the money supply, and interest rates. The focus of our analysis is on the economy's short-run fluctuations around its long-run trend.

Although there remains some debate among economists about how to analyse short-run fluctuations, and in the preceding chapter we introduced the IS-LM model, in this chapter we will look at the *model of AD and AS*, which is widely used among economists. Learning how to use this model for analysing the short-run effects of various events and policies is the primary task ahead.

THREE KEY FACTS ABOUT ECONOMIC FLUCTUATIONS

Short-run fluctuations in economic activity occur in all countries and in all times throughout history. As a starting point for understanding these year-to-year fluctuations, let's remind ourselves of some of their most important properties.

Fact 1: Economic Fluctuations are Irregular and Unpredictable

Economic fluctuations correspond to changes in business conditions. When real GDP grows rapidly, business is good. During such periods of economic expansion, firms find that customers are plentiful and that profits are growing. On the other hand, when real GDP falls during recessions, businesses have trouble. During such periods of economic contraction, many firms experience declining sales and dwindling profits. We saw in Chapter 30 that economic fluctuations are not at all regular, and they are almost impossible to predict with much accuracy.

Fact 2: Most Macroeconomic Quantities Fluctuate Together

Real GDP is the variable that is most commonly used to monitor short-run changes in the economy because it is the most comprehensive measure of economic activity. Real GDP measures the value of all final goods and services produced within a given period of time. It also measures the total income (adjusted for inflation) of everyone in the economy.

It turns out, however, that for monitoring short-run fluctuations, it does not really matter which measure of economic activity one looks at. Most macroeconomic variables that measure some type of income, spending or production, fluctuate closely together. When real GDP falls in a recession, so do personal income, corporate profits, consumer spending, investment spending, industrial production, retail sales,

home sales, auto sales and so on. Because recessions are economy-wide phenomena, they show up in many sources of macroeconomic data.

Although many macroeconomic variables fluctuate together, they fluctuate by different amounts. In particular, investment spending varies greatly over the business cycle. When economic conditions deteriorate, much of the decline is attributable to reductions in spending on new factories, housing and inventories. Investment in the UK, for example, fell by 5.1 per cent in 2008, by –11.4 per cent in 2009, and was just positive at 0.1 per cent in 2010 before another fall of 2.9 per cent in 2011.

Fact 3: As Output Falls, Unemployment Rises

Changes in the economy's output of goods and services are strongly correlated with changes in the economy's utilization of its labour force. In other words, when real GDP declines the rate of unemployment rises. The negative relationship between unemployment and real GDP is referred to as **Okun's law** after the Yale economist who published his observations in the 1960s. Okun noted that in order to keep the unemployment rate steady, real GDP needs to grow at or close to its potential. If the unemployment rate is to be reduced then real GDP must grow above potential. Being more specific, to reduce the unemployment rate by 1 per cent in a year, real GDP must rise by around 2 per cent more than potential GDP over the year.

Okun's law a 'law' which is based on observations that in order to keep the unemployment rate steady, real GDP needs to grow at or close to its potential

Okun's findings are hardly surprising: when firms choose to produce a smaller quantity of goods and services, they lay off workers, expanding the pool of unemployed. However, there is generally a time-lag between any downturn in economic activity and a rise in unemployment and vice versa. Even when positive growth resumes, therefore, unemployment is likely to continue to rise for some time afterwards. As we noted in Chapter 30, unemployment is referred to as a 'lagged indicator'.

EXPLAINING SHORT-RUN ECONOMIC FLUCTUATIONS

Describing the patterns that economies experience as they fluctuate over time is easy. Explaining what causes these fluctuations is more difficult. Indeed, compared to the topics we have studied in previous chapters, the theory of economic fluctuations remains controversial. In this chapter and the next two chapters, we develop the model that most economists use to explain short-run fluctuations in economic activity.

How the Short Run Differs from the Long Run

In previous chapters we developed theories to explain what determines most important macroeconomic variables in the long run. We have looked at the level and growth of productivity and real GDP; explained how the financial system works and how the real interest rate adjusts to balance saving and investment; why there is always some unemployment in the economy; explained the monetary system and how changes in the money supply affect the price level, the inflation rate and the nominal interest rate; and then extended this analysis to open economies in order to explain the trade balance and the exchange rate.

All of this previous analysis was based on two related ideas – the classical dichotomy and monetary neutrality. Recall that the classical dichotomy is the separation of variables into real variables (those that measure quantities or relative prices) and nominal variables (those measured in terms of money). According to classical macroeconomic theory, changes in the money supply affect nominal variables but not real variables. As a result of this monetary neutrality, we were able to examine the determinants of real

variables (real GDP, the real interest rate and unemployment) without introducing nominal variables (the money supply and the price level).

Do these assumptions of classical macroeconomic theory apply to the world in which we live? The answer to this question is of central importance to understanding how the economy works: most economists believe that classical theory describes the world in the long run but not in the short run. Beyond a period of several years, changes in the money supply affect prices and other nominal variables but do not affect real GDP, unemployment or other real variables. When studying year-to-year changes in the economy, however, the assumption of monetary neutrality is no longer appropriate. Most economists believe that, in the short run, real and nominal variables are highly intertwined. In particular, changes in the money supply can temporarily push output away from its long-run trend.

To understand the economy in the short run, therefore, we need a new model. To build this new model, we rely on many of the tools we have developed in previous chapters, but we have to abandon the classical dichotomy and the neutrality of money.

The Basic Model of Economic Fluctuations

Our model of short-run economic fluctuations focuses on the behaviour of two variables. The first variable is the economy's output of goods and services, as measured by real GDP. The second variable is the overall **price level**, an average of prices of all goods and services in an economy as measured by the CPI or the GDP deflator. Notice that output is a real variable, whereas the price level is a nominal variable. Hence, by focusing on the relationship between these two variables, we are highlighting the breakdown of the classical dichotomy.

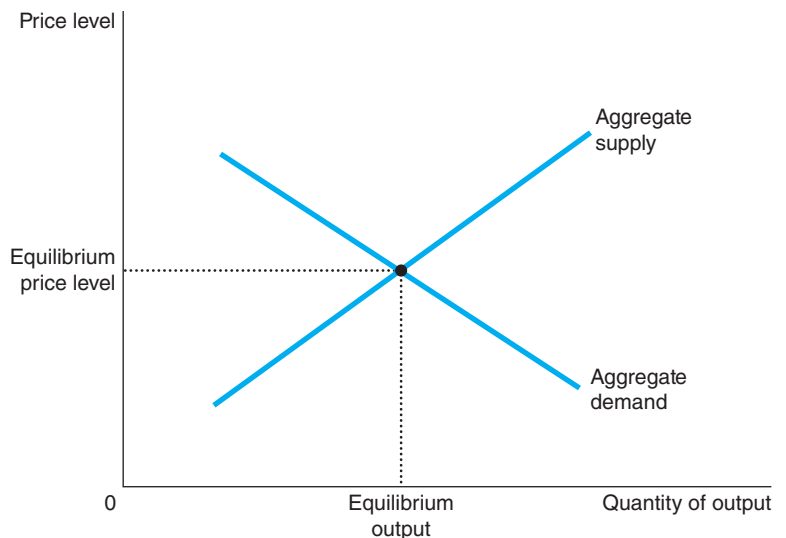
price level an average of prices of all goods and services in an economy as measured by the CPI or the GDP deflator

We analyse fluctuations in the economy as a whole with the **model of aggregate demand (AD) and aggregate supply (AS)**, which is illustrated in Figure 32.1. On the vertical axis is the overall price level in the economy. On the horizontal axis is the overall quantity of goods and services. The **aggregate demand curve**

FIGURE 32.1

Aggregate Demand and Aggregate Supply

Economists use the model of AD and AS to analyse economic fluctuations. On the vertical axis is the overall level of prices. On the horizontal axis is the economy's total output of goods and services. Output and the price level adjust to the point at which the AS and AD curves intersect.



shows the quantity of goods and services that households, firms and the government want to buy at each price level. The **aggregate supply curve** shows the quantity of goods and services that firms produce and sell at each price level. According to this model, the price level and the quantity of output adjust to bring AD and AS into balance.

model of aggregate demand and aggregate supply the model that most economists use to explain short-run fluctuations in economic activity around its long-run trend

aggregate demand curve a curve that shows the quantity of goods and services that households, firms and the government want to buy at each price level

aggregate supply curve a curve that shows the quantity of goods and services that firms choose to produce and sell at each price level

It may be tempting to view the model of AD and AS as nothing more than a large version of the model of market demand and market supply, which we introduced in Chapter 3. Yet in fact this model is quite different. When we consider demand and supply in a particular market for a good such as tablet computers, the behaviour of buyers and sellers depends on the ability of resources to move from one market to another. When the price of tablet computers rises, the quantity demanded falls because buyers will use their incomes to buy products other than tablets. Similarly, a higher price of tablets raises the quantity supplied because firms that produce these gadgets can increase production by hiring workers away from other parts of the economy. This *microeconomic* substitution from one market to another is impossible when we are analysing the economy as a whole. After all, the quantity that our model is trying to explain – real GDP – measures the total quantity produced in all of the economy's markets. To understand why the AD curve is downwards sloping and why the AS curve is upwards sloping, we need a *macroeconomic* theory. Developing such a theory is our next task.

SELF TEST How does the economy's behaviour in the short run differ from its behaviour in the long run? • Draw the model of AD and AS. What variables are on the two axes?

THE AGGREGATE DEMAND CURVE

The AD curve tells us the quantity of all goods and services demanded in the economy at any given price level. As Figure 32.2 illustrates, the AD curve is downwards sloping reflecting the inverse relationship between the price level and national income we outlined in Chapter 31. This means that, other things equal, a fall in the economy's overall level of prices (from, say, P_1 to P_2) tends to raise the quantity of goods and services demanded (from Y_1 to Y_2).

Why the Aggregate Demand Curve Slopes Downwards

Why does a fall in the price level raise the quantity of goods and services demanded? To answer this question it is useful to recall that GDP (which we denote as Y) is the sum of consumption (C), investment (I), government purchases (G) and net exports (NX):

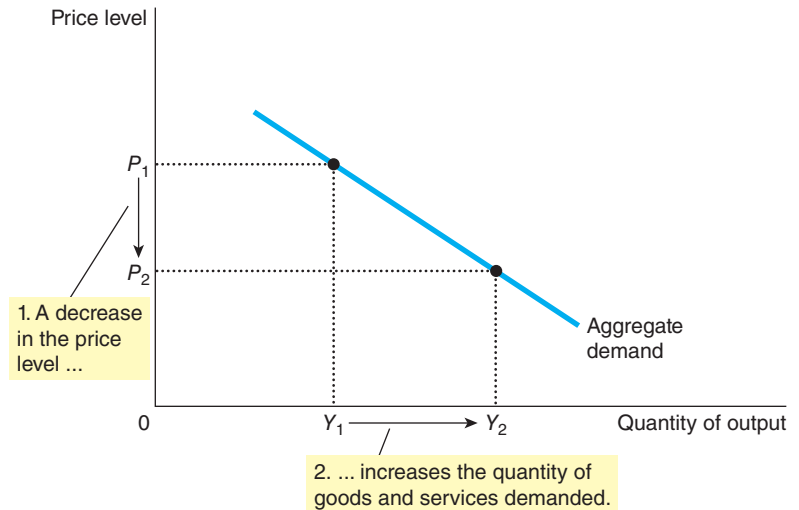
$$Y = C + I + G + NX$$

Each of these four components contributes to the AD for goods and services. For now, we assume that government spending is fixed by policy. The other three components of spending – consumption, investment and net exports – depend on economic conditions and, in particular, on the price level. To understand

the downwards slope of the AD curve, therefore, we must examine how the price level affects the quantity of goods and services demanded for consumption, investment and net exports.

FIGURE 32.2**The Aggregate Demand Curve**

A fall in the price level from P_1 to P_2 increases the quantity of goods and services demanded from Y_1 to Y_2 . There are three reasons for this negative relationship. As the price level falls, real wealth rises, interest rates fall and the exchange rate depreciates. These effects stimulate spending on consumption, investment and net exports. Increased spending on these components of output means a larger quantity of goods and services demanded.



The Price Level and Consumption: The Wealth Effect Consider the money that you hold in your pocket and your bank account. The nominal value of this money is fixed, but its real value is not. When prices fall, this money is more valuable because then it can be used to buy more goods and services. Thus, a decrease in the price level makes consumers wealthier, which in turn encourages them to spend more. The increase in consumer spending means a larger quantity of goods and services demanded.

The Price Level and Investment: The Interest Rate Effect As we discussed in Chapter 27, the price level is one determinant of the quantity of money demanded. The lower the price level, the less money households need to hold to buy the goods and services they want. When the price level falls, therefore, households try to reduce their holdings of money by lending some of it out which in turn increases the supply of real money balances. For instance, a household might use its excess money to buy interest-bearing bonds. Or it might deposit its excess money in an interest-bearing savings account, and the bank would use these funds to make more loans. In either case, as households try to convert some of their money into interest-bearing assets, they drive down interest rates. Lower interest rates, in turn, encourage borrowing by firms that want to invest in new factories and equipment and by households who want to invest in new housing. Thus, a lower price level reduces the interest rate, encourages greater spending on investment goods, and thereby increases the quantity of goods and services demanded.

The Price Level and Net Exports: The Exchange Rate Effect As we have just discussed, a lower price level lowers the interest rate. In response, some investors will seek higher returns by investing abroad. For instance, as the interest rate on European government bonds falls, an investment fund might sell European government bonds in order to buy US government bonds. As the investment fund tries to convert its euros into dollars in order to buy the US bonds, it increases the supply of euros in the market for foreign currency exchange. The increased supply of euros causes the euro to depreciate relative to other currencies. Because each euro buys fewer units of foreign currencies, non-European goods (i.e. imports) become more expensive to European residents but exporters find that foreign buyers get more euros for each unit of their currency. This change in the real exchange rate (the relative price of domestic and foreign goods) increases European exports of goods and services and decreases European imports of goods and services. Net exports, which equal exports minus imports, also increase. Thus, when a fall in the European

price level causes European interest rates to fall, the real value of the euro falls, and this depreciation stimulates European net exports and thereby increases the quantity of goods and services demanded in the European economy.

Summary There are, therefore, three distinct but related reasons why a fall in the price level increases the quantity of goods and services demanded: (1) consumers are wealthier, which stimulates the demand for consumption goods; (2) interest rates fall, which stimulates the demand for investment goods; and (3) the exchange rate depreciates, which stimulates the demand for net exports. For all three reasons, the AD curve slopes downwards.

It is important to keep in mind that the AD curve (like all demand curves) is drawn holding ‘other things equal’. In particular, our three explanations of the downwards sloping AD curve assume that the money supply is fixed. That is, we have been considering how a change in the price level affects the demand for goods and services, holding the amount of money in the economy constant. As we will see, a change in the quantity of money shifts the AD curve. At this point, just keep in mind that the AD curve is drawn for a given quantity of money.

Why the Aggregate Demand Curve Might Shift

In Chapter 31, we noted how changes in monetary and fiscal policy can cause a shift in the LM and IS curves and, at a given price level, a shift in AD. We are going to explore shifts in the AD curve in more detail here. The downwards slope of the AD curve shows that a fall in the price level raises the overall quantity of goods and services demanded. Many other factors, however, affect the quantity of goods and services demanded at a given price level. When one of these other factors changes, the AD curve shifts.

Let’s consider some examples of events that shift AD. We can categorize them according to which component of spending is most directly affected.

Shifts Arising from Consumption Suppose people suddenly become more concerned about saving for retirement and, as a result, reduce their current consumption. Because the quantity of goods and services demanded at any price level is lower, the AD curve shifts to the left. Conversely, imagine that a stock market boom makes people wealthier and less concerned about saving. The resulting increase in consumer spending means a greater quantity of goods and services demanded at any given price level, so the AD curve shifts to the right.

Thus, any event that changes how much people want to consume at a given price level shifts the AD curve. One policy variable that has this effect is the level of taxation. When the government cuts taxes, it encourages people to spend more, so the AD curve shifts to the right. When the government raises taxes, people cut back on their spending and the AD curve shifts to the left.

Shifts Arising from Investment Any event that changes how much firms want to invest at a given price level also shifts the AD curve. For instance, imagine that the computer industry introduces a faster line of computers, and many firms decide to invest in new computer systems. Because the quantity of goods and services demanded at any price level is higher, the AD curve shifts to the right. Conversely, if firms become pessimistic about future business conditions, they may cut back on investment spending, shifting the AD curve to the left.

Tax policy can also influence AD through investment. An investment tax credit (a tax rebate tied to a firm’s investment spending) increases the quantity of investment goods that firms demand at any given interest rate. It therefore shifts the AD curve to the right. The repeal of an investment tax credit reduces investment and shifts the AD curve to the left.

Another policy variable that can influence investment and AD is the money supply. As we discuss more fully in the next chapter, an increase in the money supply lowers the interest rate in the short run (the LM curve shifts to the right). This makes borrowing less costly, which stimulates investment spending and thereby shifts the AD curve to the right at a given price level. Conversely, a decrease in the money supply shifts the LM curve to the left and raises the interest rate, discourages investment spending, and

thereby shifts the AD curve to the left at the given price level. Many economists believe that changes in monetary policy have been an important source of shifts in AD in most developed economies at some points in their history.

Shifts Arising from Government Purchases The most direct way that policymakers shift the AD curve is through government purchases. For example, suppose the government decides to reduce purchases of new weapons systems. Because the quantity of goods and services demanded at any price level is lower, the AD curve shifts to the left. Conversely, if the government starts building more motorways, the result is a greater quantity of goods and services demanded at any price level, so the AD curve shifts to the right.

Shifts Arising from Net Exports Any event that changes net exports for a given price level also shifts AD. For instance, when the US experiences a recession, it buys fewer goods from Europe. This reduces European net exports and shifts the AD curve for the European economy to the left. When the US recovers from its recession, it starts buying European goods again, shifting the AD curve to the right.

Net exports sometimes change because of movements in the exchange rate. Suppose, for instance, that international speculators bid up the value of the euro in the market for foreign currency exchange. This appreciation of the euro would make goods produced in the euro area more expensive compared to foreign goods, which would depress net exports and shift the AD curve to the left. Conversely, a depreciation of the euro stimulates net exports and shifts the euro area AD curve to the right.

Summary In the next chapter we analyse the AD curve in more detail in relation to the tools of monetary and fiscal policy which can shift AD and whether policymakers should use these tools for that purpose. At this point, however, you should have some idea about why the AD curve slopes downwards and what kinds of events and policies can shift this curve. Table 32.1 summarizes what we have learned so far.

TABLE 32.1**The AD Curve: Summary****Why does the AD curve slope downward?**

- 1 *The wealth effect:* A lower price level increases real wealth, which encourages spending on consumption.
- 2 *The interest-rate effect:* A lower price level reduces the interest rate, which encourages spending on investment.
- 3 *The exchange-rate effect:* A lower price level causes the real exchange rate to depreciate, which encourages spending on net exports.

Why might the AD curve shift?

- 1 *Shifts arising from consumption:* An event that makes consumers spend more at a given price level (a tax cut, a stock market boom) shifts the AD curve to the right. An event that makes consumers spend less at a given price level (a tax hike, a stock market decline) shifts the AD curve to the left.
- 2 *Shifts arising from investment:* An event that makes firms invest more at a given price level (optimism about the future, a fall in interest rates due to an increase in the money supply) shifts the AD curve to the right. An event that makes firms invest less at a given price level (pessimism about the future, a rise in interest rates due to a decrease in the money supply) shifts the AD curve to the left.
- 3 *Shifts arising from government purchases:* An increase in government purchases of goods and services (greater spending on defence or motorway construction) shifts the AD curve to the right. A decrease in government purchases on goods and services (a cutback in defence or motorway spending) shifts the AD curve to the left.
- 4 *Shifts arising from net exports:* An event that raises spending on net exports at a given price level (a boom overseas, an exchange rate depreciation) shifts the AD curve to the right. An event that reduces spending on net exports at a given price level (a recession overseas, an exchange rate appreciation) shifts the AD curve to the left.

SELF TEST Explain the three reasons why the AD curve slopes downwards. • Give an example of an event that would shift the AD curve. Which way would this event shift the curve?

THE AGGREGATE SUPPLY CURVE

The AS curve tells us the total quantity of goods and services that firms produce and sell at any given price level. Unlike the AD curve, which is always downwards sloping, the AS curve shows a relationship that depends crucially on the time horizon being examined. In the long run, the AS curve is vertical, whereas in the short run the AS curve is upwards sloping. To understand short-run economic fluctuations, and how the short-run behaviour of the economy deviates from its long-run behaviour, we need to examine both the long-run AS curve and the short-run AS curve.

Why the Aggregate Supply Curve is Vertical in the Long Run

What determines the quantity of goods and services supplied in the long run? We implicitly answered this question earlier in the book when we analysed the process of economic growth. In the long run, an economy's production of goods and services (its real GDP) depends on its supplies of labour, capital and natural resources, and on the available technology used to turn these factors of production into goods and services. Because the price level does not affect these long-run determinants of real GDP, the long-run AS curve is vertical, as in Figure 32.3. In other words, in the long run, the economy's labour, capital, natural resources and technology determine the total quantity of goods and services supplied, and this quantity supplied is the same regardless of what the price level happens to be.

The vertical long-run AS curve is, in essence, just an application of the classical dichotomy and monetary neutrality. As we have already discussed, classical macroeconomic theory is based on the assumption that real variables do not depend on nominal variables. The long-run AS curve is consistent with this idea because it implies that the quantity of output (a real variable) does not depend on the level of prices (a nominal variable). As noted earlier, most economists believe that this principle works well when studying the economy over a period of many years, but not when studying year-to-year changes. Thus, the AS curve is vertical only in the long run.

One might wonder why supply curves for specific goods and services can be upwards sloping if the long-run AS curve is vertical. The reason is that the supply of specific goods and services depends on *relative prices* – the prices of those goods and services compared to other prices in the economy. For example, when the price of tablet computers rises, holding other prices in the economy constant, there is an incentive for suppliers of tablets to increase their production by taking labour, silicon, plastic and other inputs away from the production of other goods, such as mobile phones or laptop computers. By contrast, the economy's overall production of goods and services is limited by its labour, capital, natural resources and technology. Thus, when all prices in the economy rise together, there is no change in the overall quantity of goods and services supplied because relative prices and thus incentives have not changed.

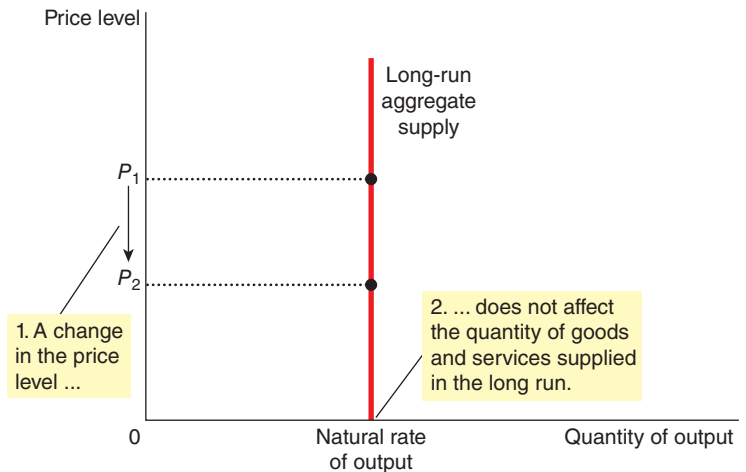
Why the Long-Run AS Curve Might Shift

The position of the long-run AS curve shows the quantity of goods and services predicted by classical macroeconomic theory. This level of production is sometimes called *potential output* or *full-employment output*. To be more accurate, we call it the **natural rate of output** because it shows what the economy produces when unemployment is at its natural, or normal, rate. The natural rate of output is the level of production towards which the economy gravitates in the long run.

natural rate of output the output level in an economy when all existing factors of production (land, labour, capital and technology resources) are fully utilized and where unemployment is at its natural rate

FIGURE 32.3**The Long-Run Aggregate Supply Curve**

In the long run, the quantity of output supplied depends on the economy's quantities of labour, capital and natural resources and on the technology for turning these inputs into output. The quantity supplied does not depend on the overall price level. As a result, the long-run AS curve is vertical at the natural rate of output.



Any change in the economy that alters the natural rate of output shifts the long-run AS curve. Because output in the classical model depends on labour, capital, natural resources and technological knowledge, we can categorize shifts in the long-run AS curve as arising from these sources.

Shifts Arising from Labour Imagine that an economy experiences an increase in immigration from abroad. Because there would be a greater number of workers, the quantity of goods and services supplied would increase. As a result, the long-run AS curve would shift to the right. Conversely, if many workers left the economy to go abroad, the long-run AS curve would shift to the left.

The position of the long-run AS curve also depends on the natural rate of unemployment, so any change in the natural rate of unemployment shifts the long-run AS curve. For example, if the government were to raise the minimum wage substantially, the natural rate of unemployment would rise, and the economy would produce a smaller quantity of goods and services. As a result, the long-run AS curve would shift to the left. Conversely, if a reform of the unemployment insurance system were to encourage unemployed workers to search harder for new jobs, the natural rate of unemployment would fall and the long-run AS curve would shift to the right.

Shifts Arising from Capital An increase in the economy's capital stock increases productivity and, thereby, the quantity of goods and services supplied. As a result, the long-run AS curve shifts to the right. Conversely, a decrease in the economy's capital stock decreases productivity and the quantity of goods and services supplied, shifting the long-run AS curve to the left.

Notice that the same logic applies regardless of whether we are discussing physical capital or human capital. An increase either in the number of machines or in the number of university degrees will raise the economy's ability to produce goods and services. Thus, either would shift the long-run AS curve to the right.

Shifts Arising from Natural Resources An economy's production depends on its natural resources, including its land, minerals and weather. A discovery of a new mineral deposit shifts the long-run AS curve to the right. A change in weather patterns that makes farming more difficult shifts the long-run AS curve to the left.

In many countries, important natural resources are imported from abroad. A change in the availability of these resources can also shift the AS curve. Events occurring in the world oil market, in particular, have historically been an important source of shifts in AS.

Shifts Arising from Technological Knowledge Perhaps the most important reason that the economy today produces more than it did a generation ago is that our technological knowledge has advanced. The invention of the computer, for instance, has allowed us to produce more goods and services from any given amounts of labour, capital and natural resources. As a result, it has shifted the long-run AS curve to the right.

Although not literally technological, there are many other events that act like changes in technology. The opening up of international trade has effects similar to inventing new production processes, so it also shifts the long-run AS curve to the right. Conversely, if the government passed new regulations preventing firms from using some production methods, perhaps because they were too dangerous for workers, the result would be a leftwards shift in the long-run AS curve.

Summary The long-run AS curve reflects the classical model of the economy we developed in previous chapters. Any policy or event that raised real GDP in previous chapters can now be viewed as increasing the quantity of goods and services supplied and shifting the long-run AS curve to the right. Any policy or event that lowered real GDP in previous chapters can now be viewed as decreasing the quantity of goods and services supplied and shifting the long-run AS curve to the left.

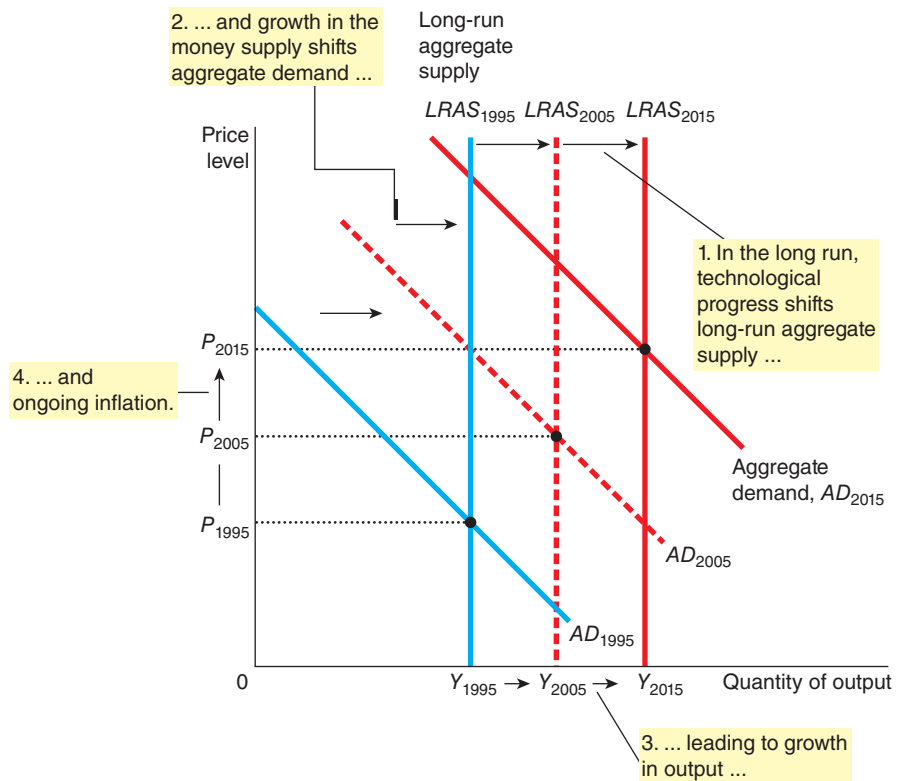
A New Way to Depict Long-Run Growth and Inflation

Having introduced the economy's AD curve and the long-run AS curve, we now have a new way to describe the economy's long-run trends. Figure 32.4 illustrates the changes that occur in the economy from decade to decade. Notice that both curves are shifting. Although there are many forces that govern the economy in the long run and can in principle cause such shifts, the two most important in practice are technology and monetary policy. Technological progress enhances the economy's ability to produce goods and services, and this continually shifts the long-run AS curve to the right. At the same time, because the central bank increases the money supply over time, the AD curve also shifts to the right. As the figure illustrates, the result is trend growth in output (as shown by increasing Y) and continuing inflation (as shown by increasing P). This is just another way of representing the classical analysis of growth and inflation we conducted earlier in the book.

FIGURE 32.4

Long-Run Growth and Inflation in the Model of Aggregate Demand and Aggregate Supply

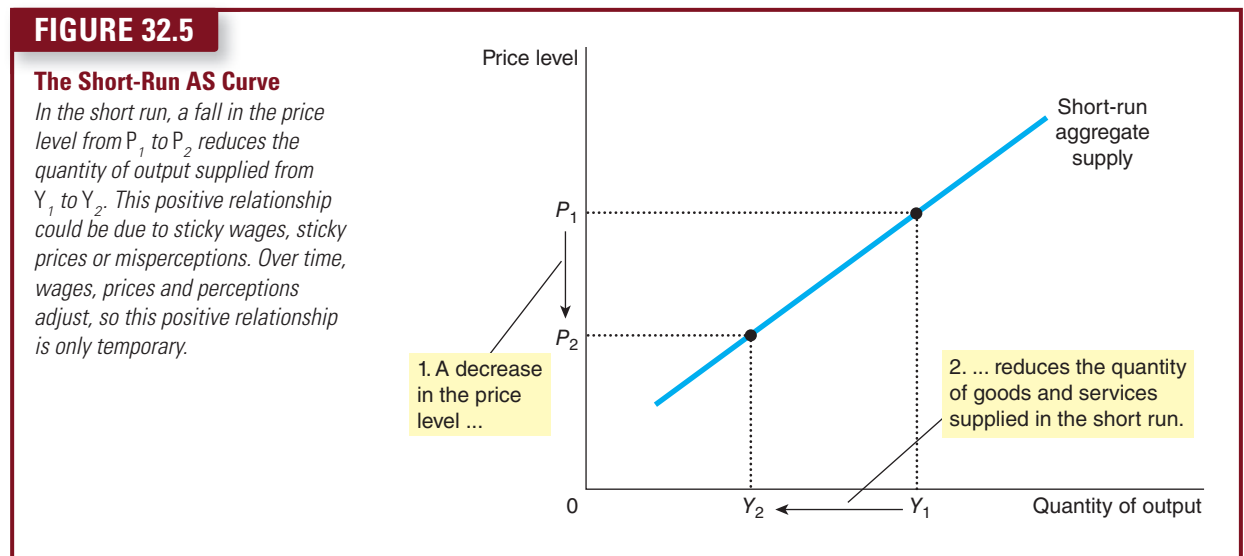
As the economy becomes better able to produce goods and services over time, primarily because of technological progress, the long-run AS curve shifts to the right. At the same time, as the central bank increases the money supply, the AD curve also shifts to the right. In this figure, output grows from Y_{1995} to Y_{2005} and then to Y_{2015} , and the price level rises from P_{1995} to P_{2005} and then to P_{2015} . Thus, the model of AD and AS offers a new way to describe the classical analysis of growth and inflation.



The purpose of developing the model of AD and AS, however, is not to dress our long-run conclusions in new clothing. Instead, it is to provide a framework for short-run analysis, as we will see in a moment. As we develop the short-run model, we keep the analysis simple by not showing the continuing growth and inflation depicted in Figure 32.4. But always remember that long-run trends provide the background for short-run fluctuations. Short-run fluctuations in output and the price level should be viewed as deviations from the continuing long-run trends.

Why the AS Curve Slopes Upward in the Short Run

We now come to the key difference between the economy in the short run and in the long run: the behaviour of AS. As we have already discussed, the long-run AS curve is vertical. By contrast, in the short run, the AS curve is upwards sloping, as shown in Figure 32.5. That is, over a period of a year or two, an increase in the overall level of prices in the economy tends to raise the quantity of goods and services supplied, and a decrease in the level of prices tends to reduce the quantity of goods and services supplied.



What causes this positive relationship between the price level and output? Macroeconomists have proposed three theories for the upward slope of the short-run AS curve. In each theory, a specific market imperfection causes the supply side of the economy to behave differently in the short run than it does in the long run. Although each of the following theories will differ in detail, they share a common theme: the quantity of output supplied deviates from its long-run or 'natural' level when the price level deviates from the price level that people *expected* to prevail. When the price level rises above the expected level, output rises above its natural rate, and when the price level falls below the expected level, output falls below its natural rate.

The Sticky Wage Theory The first and simplest explanation of the upwards slope of the short-run AS curve is the sticky wage theory which we initially encountered in Chapter 30. According to this theory, the short-run AS curve slopes upwards because nominal wages are slow to adjust, or are 'sticky', in the short run. To some extent, the slow adjustment of nominal wages is attributable to long-term contracts between workers and firms that fix nominal wages, sometimes for as long as three years. In addition, this slow adjustment may be attributable to social norms and notions of fairness that influence wage setting and that change only slowly over time.

To see what sticky nominal wages mean for AS, imagine that a firm has agreed in advance to pay its workers a certain nominal wage based on what it expected the price level to be. If the price level P falls below the level that was expected and the nominal wage remains stuck at W , then the real wage $\frac{W}{P}$ rises above the level the firm planned to pay. Because wages are a large part of a firm's production costs,

a higher real wage means that the firm's real costs have risen. The firm responds to these higher costs by hiring less labour and producing a smaller quantity of goods and services. In other words, because wages do not adjust immediately to the price level, a lower price level makes employment and production less profitable, so firms reduce the quantity of goods and services they supply.

The Sticky Price Theory As we just discussed, the sticky wage theory emphasizes that nominal wages adjust slowly over time. The sticky price theory emphasizes that the prices of some goods and services also adjust sluggishly in response to changing economic conditions. This slow adjustment of prices occurs in part because of the *menu costs* to adjusting prices. These menu costs include the cost of printing and distributing price lists or mail-order catalogues and the time required to change price tags. As a result of these costs, prices as well as wages may be sticky in the short run.

To see the implications of sticky prices for AS, suppose that each firm in the economy announces its prices in advance based on the economic conditions it expects to prevail. Then, after prices are announced, the economy experiences an unexpected contraction in the money supply, which (as we have learned) will reduce the overall price level in the long run. Although some firms reduce their prices immediately in response to changing economic conditions, other firms may not want to incur additional menu costs and, therefore, may temporarily lag behind. Because these lagging firms have prices that are too high, their sales decline. Declining sales, in turn, cause these firms to cut back on production and employment. In other words, because not all prices adjust instantly to changing conditions, an unexpected fall in the price level leaves some firms with higher-than-desired prices, and these higher-than-desired prices depress sales and induce firms to reduce the quantity of goods and services they produce.

The Misperceptions Theory A third approach to the short-run AS curve is the misperceptions theory. According to this theory, changes in the overall price level can temporarily mislead suppliers about what is happening in the individual markets in which they sell their output. As a result of these short-run misperceptions, suppliers respond to changes in the level of prices, and this response leads to an upwards sloping AS curve.

To see how this might work, suppose the overall price level falls below the level that people expected. When suppliers see the prices of their products fall, they may mistakenly believe that their *relative* prices have fallen. For example, wheat farmers may notice a fall in the price of wheat before they notice a fall in the prices of the many items they buy as consumers. They may infer from this observation that the reward to producing wheat is temporarily low, and they may respond by reducing the quantity of wheat they supply. Similarly, workers may notice a fall in their nominal wages before they notice a fall in the prices of the goods they buy. They may infer that the reward to working is temporarily low and respond by reducing the quantity of labour they supply. In both cases, a lower price level causes misperceptions about relative prices, and these misperceptions induce suppliers to respond to the lower price level by decreasing the quantity of goods and services supplied.

Summary There are three alternative explanations for the upwards slope of the short-run AS curve: (1) sticky wages; (2) sticky prices; and (3) misperceptions. Economists debate which of these theories is correct, and it is very possible each contains an element of truth. For our purposes in this book, the similarities of the theories are more important than the differences. All three theories suggest that output deviates from its natural rate when the price level deviates from the price level that people expected. We can express this mathematically as follows:

$$\text{Quantity of output supplied} = \text{Natural rate of output} + a \left(\text{Actual price level} - \text{Expected price level} \right)$$

where a is a number that determines how much output responds to unexpected changes in the price level.

Notice that each of the three theories of short-run AS emphasizes a problem that is likely to be only temporary. Whether the upwards slope of the AS curve is attributable to sticky wages, sticky prices or misperceptions, these conditions will not persist forever. Eventually, as people adjust their expectations, nominal wages adjust, prices become unstuck and misperceptions are corrected. In other words, the expected and actual price levels are equal in the long run, and the AS curve is vertical rather than upwards sloping.

Why the Short-Run AS Curve Might Shift

The short-run AS curve tells us the quantity of goods and services supplied in the short run for any given level of prices. We can think of this curve as similar to the long-run AS curve but made upwards sloping by the presence of sticky wages, sticky prices and misperceptions. Thus, when thinking about what shifts the short-run AS curve, we have to consider all those variables that shift the long-run AS curve plus a new variable – the expected price level – that influences sticky wages, sticky prices and misperceptions.

Let's start with what we know about the long-run AS curve. As we discussed earlier, shifts in the long-run AS curve normally arise from changes in labour, capital, natural resources or technological knowledge. These same variables shift the short-run AS curve. For example, when an increase in the economy's capital stock increases productivity, both the long-run and short-run AS curves shift to the right. When an increase in the minimum wage raises the natural rate of unemployment, both the long-run and short-run AS curves shift to the left.

The important new variable that affects the position of the short-run AS curve is people's expectation of the price level. As we have discussed, the quantity of goods and services supplied depends, in the short run, on sticky wages, sticky prices and misperceptions. Yet wages, prices and perceptions are set on the basis of expectations of the price level. So when expectations change, the short-run AS curve shifts.

To make this idea more concrete, let's consider a specific theory of AS – the sticky wage theory. According to this theory, when workers and firms expect the price level to be high, they are more likely to negotiate high nominal wages. High wages raise firms' costs and, for any given actual price level, reduce the quantity of goods and services that firms supply. Thus, when the expected price level rises, wages are higher, costs increase, and firms supply a smaller quantity of goods and services at any given actual price level. Thus, the short-run AS curve shifts to the left. Conversely, when the expected price level falls, wages are lower, costs decline, firms increase production at any given price level, and the short-run AS curve shifts to the right.

A similar logic applies in each theory of AS. The general lesson is the following:

- An increase in the expected price level reduces the quantity of goods and services supplied and shifts the short-run AS curve to the left.
- A decrease in the expected price level raises the quantity of goods and services supplied and shifts the short-run AS curve to the right.

As we will see in the next section, this influence of expectations on the position of the short-run AS curve plays a key role in reconciling the economy's behaviour in the short run with its behaviour in the long run. In the short run, expectations are fixed, and the economy finds itself at the intersection of the AD curve and the short-run AS curve. In the long run, expectations adjust, and the short-run AS curve shifts. This shift ensures that the economy eventually finds itself at the intersection of the AD curve and the long-run AS curve.

You should now have some understanding about why the short-run AS curve slopes upward and what events and policies can cause this curve to shift. Table 32.2 summarizes our discussion.

SELF TEST Explain why the long-run AS curve is vertical. ● Explain three theories for why the short-run AS curve is upwards sloping.

TWO CAUSES OF ECONOMIC FLUCTUATIONS

Now that we have introduced the model of AD and AS, we have the basic tools we need to analyse fluctuations in economic activity. In particular, we can use what we have learned about AD and AS to examine the two basic causes of short-run fluctuations.

To keep things simple, we assume the economy begins in long-run equilibrium, as shown in Figure 32.6. Equilibrium output and the price level are determined by the intersection of the AD curve and the long-run

AS curve, shown as point A in the figure. At this point, output is at its natural rate. The short-run AS curve passes through this point as well, indicating that wages, prices and perceptions have fully adjusted to this long-run equilibrium. That is, when an economy is in its long-run equilibrium, wages, prices and perceptions must have adjusted so that the intersection of AD with short-run AS is the same as the intersection of AD with long-run AS.

TABLE 32.2**The Short-Run Aggregate Supply Curve: Summary****Why does the short-run aggregate supply curve slope upward?**

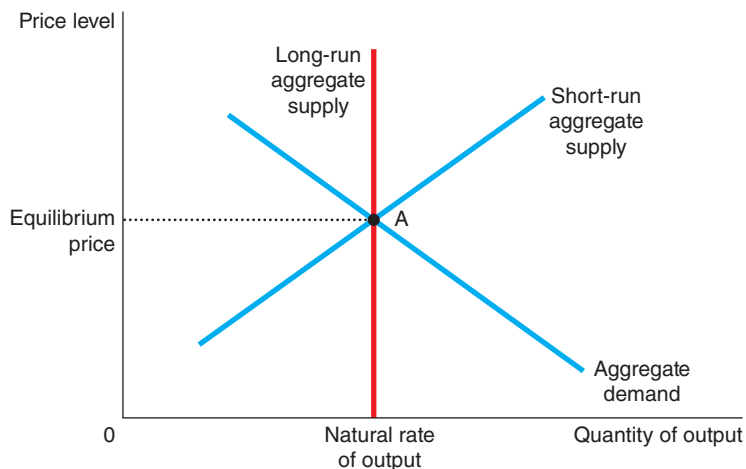
- 1 **The sticky wage theory:** An unexpectedly low price level raises the real wage, which causes firms to hire fewer workers and produce a smaller quantity of goods and services.
- 2 **The sticky price theory:** An unexpectedly low price level leaves some firms with higher-than-desired prices, which depresses their sales and leads them to cut back production.
- 3 **The misperceptions theory:** An unexpectedly low price level leads some suppliers to think their relative prices have fallen, which induces a fall in production.

Why might the short-run aggregate supply curve shift?

- 1 **Shifts arising from labour:** An increase in the quantity of labour available (perhaps due to a fall in the natural rate of unemployment) shifts the AS curve to the right. A decrease in the quantity of labour available (perhaps due to a rise in the natural rate of unemployment) shifts the AS curve to the left.
- 2 **Shifts arising from capital:** An increase in physical or human capital shifts the AS curve to the right. A decrease in physical or human capital shifts the AS curve to the left.
- 3 **Shifts arising from natural resources:** An increase in the availability of natural resources shifts the AS curve to the right. A decrease in the availability of natural resources shifts the AS curve to the left.
- 4 **Shifts arising from technology:** An advance in technological knowledge shifts the AS curve to the right. A decrease in the available technology (perhaps due to government regulation) shifts the AS curve to the left.
- 5 **Shifts arising from the expected price level:** A decrease in the expected price level shifts the short-run AS curve to the right. An increase in the expected price level shifts the short-run AS curve to the left.

FIGURE 32.6**The Long-Run Equilibrium**

The long-run equilibrium of the economy is found where the AD curve crosses the long-run AS curve (point A). When the economy reaches this long-run equilibrium, wages, prices and perceptions will have adjusted so that the short-run AS curve crosses this point as well.



The Effects of a Shift in Aggregate Demand

Suppose that for some reason a wave of pessimism suddenly overtakes the economy. The cause might be a government scandal, a crash on the stock market or the outbreak of war overseas. Because of this event, many people lose confidence in the future and alter their plans. Households cut back on their spending and delay major purchases, and firms put off buying new equipment.

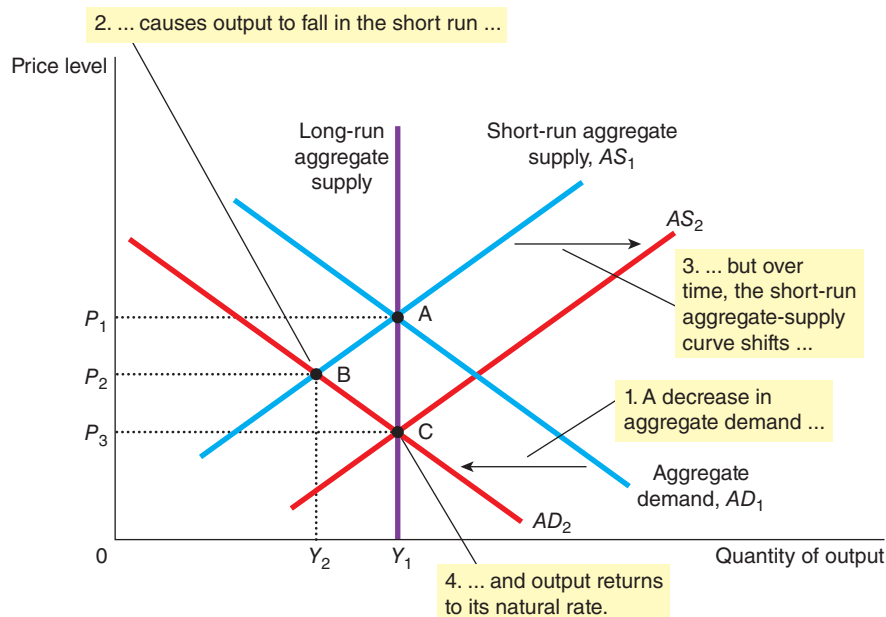
What is the impact of such a wave of pessimism on the economy? Such an event reduces the AD for goods and services. That is, for any given price level, households and firms now want to buy a smaller quantity of goods and services. As Figure 32.7 shows, the AD curve shifts to the left from AD_1 to AD_2 .

In this figure we can examine the effects of the fall in AD. In the short run, the economy moves along the initial short-run AS curve AS_1 , going from point A to point B. As the economy moves from point A to point B, output falls from Y_1 to Y_2 , and the price level falls from P_1 to P_2 . The falling level of output indicates that the economy is in a recession. Although not shown in the figure, firms respond to lower sales and production by reducing employment. Thus, the pessimism that caused the shift in AD is, to some extent, self-fulfilling: pessimism about the future leads to falling incomes and rising unemployment.

FIGURE 32.7

A Contraction in Aggregate Demand

A fall in AD, which might be due to a wave of pessimism in the economy, is represented with a leftward shift in the AD curve from AD_1 to AD_2 . The economy moves from point A to point B. Output falls from Y_1 to Y_2 , and the price level falls from P_1 to P_2 . Over time, as wages, prices and perceptions adjust, the short-run AS curve shifts to the right from AS_1 to AS_2 , and the economy reaches point C, where the new AD curve crosses the long-run AS curve. The price level falls to P_3 , and output returns to its natural rate, Y_1 .



What should policymakers do when faced with such a recession? One possibility is to take action to increase AD. As we noted earlier, an increase in government spending or an increase in the money supply would increase the quantity of goods and services demanded at any price and, therefore, would shift the AD curve to the right. If policymakers can act with sufficient speed and precision, they can offset the initial shift in AD, return the AD curve back to AD_1 and bring the economy back to point A. (The next chapter discusses in more detail the ways in which monetary and fiscal policy influence AD, as well as some of the practical difficulties in using these policy instruments.)

Even without action by policymakers, the recession will remedy itself over a period of time. Because of the reduction in AD, the price level falls. Eventually, expectations catch up with this new reality, and the expected price level falls as well. Because the fall in the expected price level alters wages, prices and perceptions, it shifts the short-run AS curve to the right from AS_1 to AS_2 in Figure 32.7. This adjustment of expectations allows the economy over time to approach point C, where the new AD curve (AD_2) crosses the long-run AS curve.

In the new long-run equilibrium, point C, output is back to its natural rate. Even though the wave of pessimism has reduced AD, the price level has fallen sufficiently (to P_3) to offset the shift in the AD curve. Thus, in the long run, the shift in AD is reflected fully in the price level and not at all in the level of output. In other words, the long-run effect of a shift in AD is a nominal change (the price level is lower) but not a real change (output is the same).

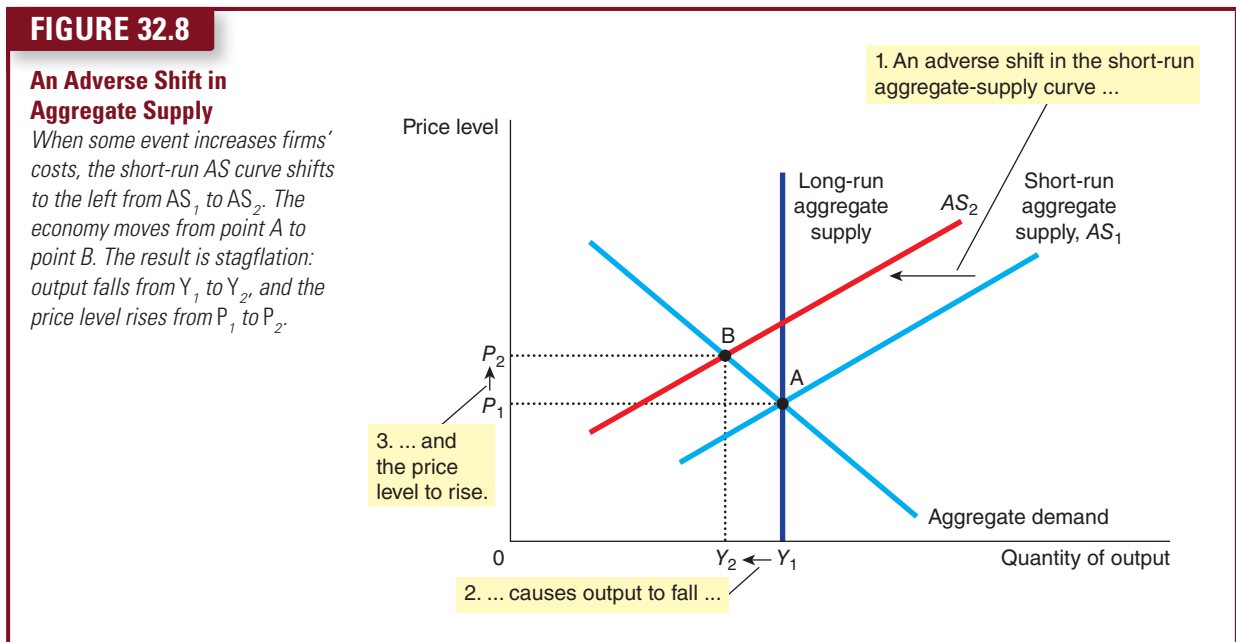
To sum up, this story about shifts in AD has two important lessons:

- In the short run, shifts in AD cause fluctuations in the economy's output of goods and services.
- In the long run, shifts in AD affect the overall price level but do not affect output.

The Effects of a Shift in Aggregate Supply

Imagine once again an economy in its long-run equilibrium. Now suppose that suddenly some firms experience an increase in their costs of production. For example, bad weather might destroy some agricultural crops, driving up the cost of producing food products. Or political or military conflict might interrupt the shipping of crude oil, driving up the cost of producing oil products.

What is the macroeconomic impact of such an increase in production costs? For any given price level, firms now want to supply a smaller quantity of goods and services. Thus, as Figure 32.8 shows, the short-run AS curve shifts to the left from AS_1 to AS_2 . (Depending on the event, the long-run AS curve might also shift. To keep things simple, however, we will assume that it does not.)



In this figure we can trace the effects of the leftward shift in AS. In the short run, the economy moves along the existing AD curve, going from point A to point B. The output of the economy falls from Y_1 to Y_2 and the price level rises from P_1 to P_2 . Because the economy is experiencing both *stagnation* (falling output) and *inflation* (rising prices), such an event is sometimes called **stagflation**.

stagflation a period of falling output and rising prices

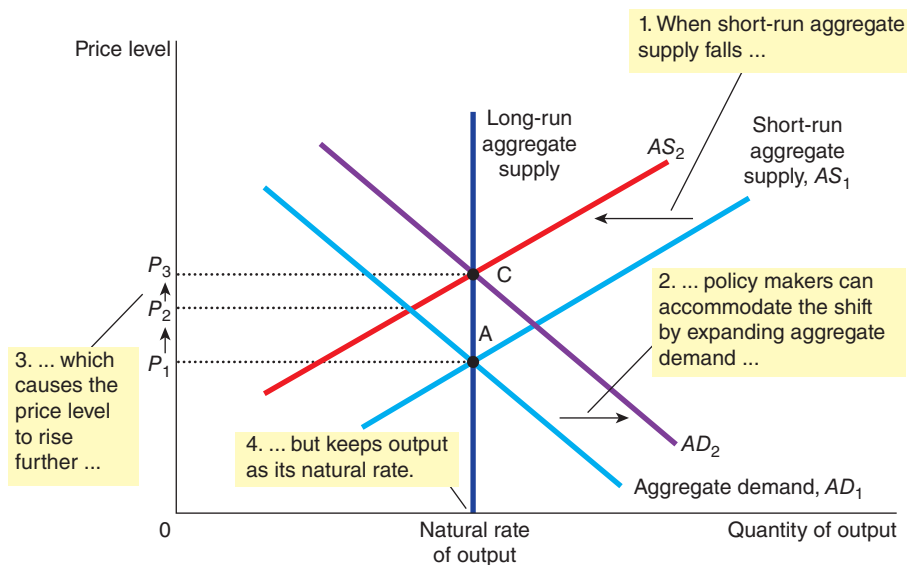
What should policymakers do when faced with stagflation? There are no easy choices. One possibility is to do nothing. In this case, the output of goods and services remains depressed at Y_2 for a while. Eventually, however, the recession will remedy itself as wages, prices and perceptions adjust to raise production costs. A period of low output and high unemployment, for instance, puts downwards pressure on workers' wages. Lower wages, in turn, increase the quantity of output supplied. Over time, as the short-run AS curve shifts back towards AS_1 , the price level falls, and the quantity of output approaches its natural rate. In the long run, the economy returns to point A, where the AD curve crosses the long-run AS curve. This is the view that believers of free markets might adopt.

Alternatively, policymakers who control monetary and fiscal policy might attempt to offset some of the effects of the shift in the short-run AS curve by shifting the AD curve. This possibility is shown in Figure 32.9. In this case, changes in policy shift the AD curve to the right from AD_1 to AD_2 – exactly enough to prevent the shift in AS from affecting output. The economy moves directly from point A to point C. Output remains at its natural rate, and the price level rises from P_1 to P_3 . In this case, policymakers are said to *accommodate* the shift in AS because they allow the increase in costs to permanently affect the level of prices. This intervention by policymakers would be seen as being desirable by supporters of Keynes.

FIGURE 32.9

Accommodating an Adverse Shift in Aggregate Supply

Faced with an adverse shift in AS from AS_1 to AS_2 , policymakers who can influence AD might try to shift the AD curve to the right from AD_1 to AD_2 . The economy would move from point A to point C. This policy would prevent the supply shift from reducing output in the short run, but the price level would permanently rise from P_1 to P_3 .



These different views on policy action form a key aspect of the debate between economists about action in the face of short-run fluctuations in economic activity.

To sum up, this story about shifts in AS has two important lessons:

- Shifts in AS can cause stagflation – a combination of recession (falling output) and inflation (rising prices).
- Policymakers who can influence AD cannot offset both of these adverse effects simultaneously.

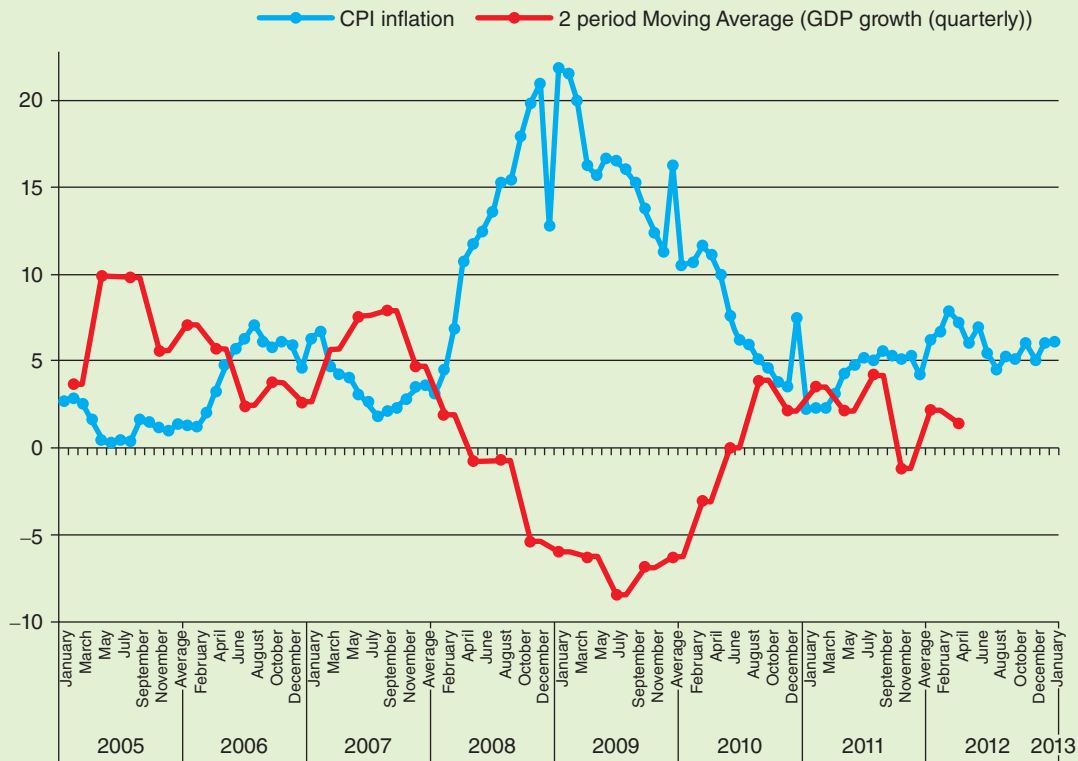
CASE STUDY Stagflation in Iceland?

The economy of Iceland was badly affected by the financial crisis with three of its main banks, Kaupthing, Landsbanki and Glitnir, all failing. The collapse of these banks led to the Icelandic economy slumping. Figure 32.10 shows how significant and rapid the decline in the economy was between mid-2007 and mid-2009. The annual change in GDP was reported as -18.2 per cent in 2008 and -9.1 per cent in 2009. The economy began to improve slowly and returned to growth in early 2010 and growing at a rate of 3–4 per cent between the middle of 2011 to November 2012 when the economy contracted once again.

FIGURE 32.10

Inflation and Economic Growth in Iceland

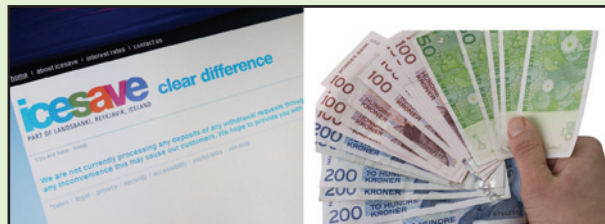
The figure charts inflation and economic growth measured as a change from the previous quarter between 2005 and 2012.



Inflation, meanwhile, has been a problem. Between 2005 and March 2006, inflation was under 2.0 per cent but began to accelerate after that time before easing in the latter part of 2007. Once the full effects of the financial crisis took hold, inflation accelerated, quickly reaching a high of 21.9 per cent in January 2009. Inflation has eased since then but the trend since November 2010 is upwards.

The aftermath of the financial crisis left Iceland facing the twin problems of accelerating inflation and shrinking economic activity. It can be argued that the particular

circumstances of the banking collapse in Iceland led to the stagflation but what of the future for Iceland? There are fears that Iceland could again be heading for a period of stagflation. The rate of growth of inflation seems to be accelerating and some analysts expect inflation to be over 5 per cent in the coming years but at the same time growth seems to be slowing. Statistics Iceland reported that preliminary GDP figures for 2012 would show growth of 1.6 per cent – not a contraction in the economy but a growth rate below that forecast. There was a 50 per cent chance that these figures could be revised downwards.



The collapse of the banking system in Iceland not only affected the people of Iceland but across many other parts of Europe too.

SELF TEST Suppose that the election of a popular prime minister suddenly increases people's confidence in the future. Use the model of AD and AS to analyse the effect on the economy.

NEW KEYNESIAN ECONOMICS

We introduced Keynesian economics and the IS-LM model in Chapter 31. Recall that Keynesian economics developed in response to the depression of the 1930s, when it appeared that the assumptions of classical economics that markets would clear, were failing. Keynesian ideas were the mainstay for economic policy across the developed world in the post-war era and there was something of a consensus amongst economists that our understanding of the macroeconomy was considerable. The latter part of the 1960s began to reveal some flaws in that assumption as global conditions changed. The breakdown of fixed exchange rates in the late 1960s and the oil crisis in the early 1970s presented macroeconomists with significant challenges in explaining the economic conditions that existed in some countries – stagflation in particular.

Economics came to be classified as a debate between Keynesians on the one hand, who pointed to markets not clearing quickly, and the neo or new classicists on the other hand, who re-emphasized the efficiency of markets and argued that microeconomics provided a foundation for understanding macroeconomics.

Out of this debate emerged a group who were referred to as the New Keynesians. New Keynesian economics placed an emphasis on providing sound microeconomic principles to underpin Keynesian macroeconomics. Specifically, the New Keynesians sought to explain how price and wage stickiness had its foundation in the microeconomic analysis of labour markets and price setting by firms.

Features of New Keynesian Economics

New Keynesians would argue that short-run fluctuations in economic output violate the classical dichotomy, as outlined earlier in this chapter. Changes in nominal variables like the money supply do have an influence on output and employment – real variables. In addition, they would argue that to develop an understanding of changes in economic activity, an understanding of the imperfections in the economy is necessary – that firms operate under imperfect competition, that consumers and firms are subject to imperfect information and that there are built-in rigidities in the economy which hinder the movement of prices and wages as we have noted.

So, can we conclude from this that the economists classed as New Keynesians are supporters of demand management but with some reservations? The answer to this is 'no'. The splitting of economics into different schools of thought has been convenient for the mass media to present disagreements between economists in relatively simple terms. The reality is that there is probably more agreement in economics that most members of the public would imagine. A difference in some base assumptions and a difference in the relative size of parameters in models is largely the reason why economists disagree. Many economists would subscribe to the view that changes in the money supply affect aggregate demand (a so-called 'monetarist' view) but also agree that price rigidities in the macroeconomy exist because of imperfections in the microeconomy. Some New Keynesian economists would, equally, not be in support of policy intervention to correct short-term fluctuations in the economy by attempting to manipulate aggregate demand. The main reason being that they would argue that the effects of such policy interventions take time to work through the economy by which time economic conditions have changed and so the policy will not be as effective as anticipated. Changes in interest rates, for example, are estimated to take around 18 months to work through the economy; tax changes might be received in different ways by different people and it is never certain enough to predict how people will react to a tax cut or a tax increase in any precise quantifiable way to make such a policy option a clear-cut choice. Equally, policies to increase government spending, on infrastructure projects for example, often take many years to have any effect even if such projects are 'shovel ready' – able to be put in operation quickly.

The response to the financial crisis has reinvigorated the debate between whether active intervention by central banks and governments to boost the economy in the face of persistent sluggish growth and

rising unemployment is the best policy response, or whether a 'do nothing' option is preferable to allow markets to adjust to changed conditions. It is always easy to seize on one headline statement by any economist and immediately seek to pigeon-hole that view. New Keynesian economics reminds us that the complexities of the economy perhaps deserve a more discerning look at the detail behind what economists are actually saying and how much they agree and why they agree, as much as why they might disagree and the extent of such disagreements.

CONCLUSION

This chapter has achieved two goals. First, we have discussed some of the important facts about short-run fluctuations in economic activity. Second, we have introduced a basic model to explain those fluctuations, called the model of AD and AS. We continue our study of this model in the next chapter in order to understand more fully what causes fluctuations in the economy and how policymakers might respond to these fluctuations.

IN THE NEWS



Expectations

One of the key tasks of policymakers is to set the tone for expectations in the economy – such expectations have to be realistic and also believed by the population if the policy is to have any credence.

Moderating Inflationary Expectations in China

The Governor of the central bank of China, the People's Bank of China (PBoC), Zhou Xiaochuan, re-emphasized the commitment of the PBoC to keeping inflation under control in a speech made in March 2013. The comments followed the release of inflation data which showed the CPI accelerating to 3.2 per cent in February 2013, a ten-month high. Zhou noted that this was higher than forecast and that this emphasized the need to ensure the PBoC acted to moderate inflationary expectations in the economy.

To reinforce the message that inflation would be kept under control, Zhou announced that the growth in the money supply, as measured by M2,

would be tightened with a target of 13 per cent growth in M2 for 2013. Zhou noted that this target was lower than the actual growth in M2 in 2012, which was over 14 per cent and reflected the desire by the PBoC to help boost economic growth in China. Part of the pro-growth policy that was a focus of 2012 included government spending on infrastructure and the looser monetary policy also allowed more people to be able to borrow money to buy property. However, the PBoC noted that house prices increased over 12 per cent in Beijing and had risen in 53 of 70 major cities across the country and house price inflation was something that it wanted to control.

Whilst China intends to tighten monetary policy, the PBoC also noted that it was monitoring the potential effect on China's economy of looser

monetary policy being adopted elsewhere in the world. Quantitative easing in the US and UK should lead to a higher money supply and was likely to lead to a weakening of the pound and dollar against the yuan. A stronger yuan would affect China's exporters and also raise the cost of imports which would put pressure on inflation.

Questions

- 1 Using your understanding of the model of AD and AS, what factors might be causing an increase in inflation in China?
- 2 Why might the increase in the rate of inflation in March 2013 to a ten-month high, trigger expectations of higher inflation in the future and what effect might this have on the Chinese economy in the absence of policy changes by the Chinese authorities?
- 3 Why would the PBoC's announcement that monetary policy be tightened influence expectations and what would you think the results of such a policy announcement, assuming it was enforced, would be?
- 4 Assuming the growth of M2 is tightened, use the AD and AS model to analyse the likely effect on inflation and national income in China.
- 5 Why would a loose monetary policy in countries like the US and UK lead to a strengthening of the Chinese currency and what would you expect the effect on AD in China to be as a result?

New housing development in China is partly in response to rising wage levels fuelling increased demand which in turn is leading to rising house prices.



SUMMARY

- All societies experience short-run economic fluctuations around long-run trends. These fluctuations are irregular and largely unpredictable. When recessions do occur, real GDP and other measures of income, spending and production fall, and unemployment rises.
- Economists analyse short-run economic fluctuations using the model of AD and AS. According to this model, the output of goods and services and the overall level of prices adjust to balance AD and AS.
- The AD curve slopes downwards for three reasons. First, a lower price level raises the real value of households' money holdings, which stimulates consumer spending. Second, a lower price level reduces the quantity of money households' demand; as households try to convert money into interest-bearing assets, interest rates fall, which stimulates investment spending. Third, as a lower price level reduces interest rates, the local currency depreciates in the market for foreign currency exchange, which stimulates net exports.
- Any event or policy that raises consumption, investment, government purchases or net exports at a given price level increases AD. Any event or policy that reduces consumption, investment, government purchases or net exports at a given price level decreases AD.
- The long-run AS curve is vertical. In the long run, the quantity of goods and services supplied depends on the economy's labour, capital, natural resources and technology, but not on the overall level of prices.
- Three theories have been proposed to explain the upwards slope of the short-run AS curve. According to the sticky wage theory, an unexpected fall in the price level temporarily raises real wages, which induces firms to reduce employment and production. According to the sticky price theory, an unexpected fall in the price level leaves some firms with prices that are temporarily too high, which reduces their sales and causes them to cut back production. According to the misperceptions theory, an unexpected fall in the price level leads suppliers to mistakenly believe that their relative prices have fallen, which induces them to reduce production. All three theories imply that output deviates from its natural rate when the price level deviates from the price level that people expected.
- Events that alter the economy's ability to produce output, such as changes in labour, capital, natural resources or technology, shift the short-run AS curve (and may shift the long-run AS curve as well). In addition, the position of the short-run AS curve depends on the expected price level.
- One possible cause of economic fluctuations is a shift in AD. When the AD curve shifts to the left, for instance, output and prices fall in the short run. Over time, as a change in the expected price level causes wages, prices and perceptions to adjust, the short-run AS curve shifts to the right, and the economy returns to its natural rate of output at a new, lower price level.
- A second possible cause of economic fluctuations is a shift in AS. When the AS curve shifts to the left, the short-run effect is falling output and rising prices – a combination called stagflation. Over time, as wages, prices and perceptions adjust, the price level falls back to its original level, and output recovers.
- New Keynesian economics represents a research tradition that questions the classical dichotomy and recognizes imperfections in the economy as key elements in explaining short run deviations from trend.

QUESTIONS FOR REVIEW

- 1 What do you understand by the term 'economic activity'?
- 2 What is the official definition of a recession?
- 3 Name two macroeconomic variables that decline when the economy goes into a recession. Name one macroeconomic variable that rises during a recession.
- 4 Draw a diagram with aggregate demand, short-run aggregate supply and long-run aggregate supply. Be careful to label the axes correctly.
- 5 List and explain the three reasons why the aggregate demand curve is downwards sloping.
- 6 Explain why the long-run aggregate supply curve is vertical.
- 7 List and explain the three theories for why the short-run aggregate supply curve is upwards sloping.
- 8 What might shift the aggregate demand curve to the left? Use the model of aggregate demand and aggregate supply to trace through the effects of such a shift.
- 9 What might shift the aggregate supply curve to the left? Use the model of aggregate demand and aggregate supply to trace through the effects of such a shift.
- 10 What outcomes are possible if both aggregate demand and aggregate supply are both shifting? Use the model of aggregate demand and aggregate supply to outline some of these possible outcomes.

PROBLEMS AND APPLICATIONS

- 1 Why do you think that investment is more variable over the business cycle than consumer spending? Which category of consumer spending do you think would be most volatile: durable goods (such as furniture and car purchases), non-durable goods (such as food and clothing) or services (such as haircuts and medical care)? Why?
- 2 Suppose that the economy is in a long-run equilibrium.
 - a. Use a diagram to illustrate the state of the economy. Be sure to show AD, short-run AS and long-run AS.
 - b. Now suppose that a financial crisis causes AD to fall. Use your diagram to show what happens to output and the price level in the short run. What happens to the unemployment rate?
 - c. Use the sticky wage theory of AS to explain what will happen to output and the price level in the long run (assuming there is no change in policy). What role does the expected price level play in this adjustment? Be sure to illustrate your analysis with a graph.
- 3 Explain whether each of the following events will increase, decrease or have no effect on long-run AS.
 - a. The country experiences a wave of immigration.
 - b. The government raises the minimum wage above the national average wage level.
 - c. A war leads to the destruction of a large number of factories.
- 4 In Figure 32.7, how does the unemployment rate at points B and C compare to the unemployment rate at point A? Under the sticky wage explanation of the short-run AS curve, how does the real wage at points B and C compare to the real wage at point A?
- 5 Explain why the following statements are false.
 - a. 'The AD curve slopes downwards because it is the horizontal sum of the demand curves for individual goods.'
 - b. 'The long-run AS curve is vertical because economic forces do not affect long-run AS.'
 - c. 'If firms adjusted their prices every day, then the short run AS curve would be horizontal.'
 - d. 'Whenever the economy enters a recession, its long-run AS curve shifts to the left.'
- 6 For each of the three theories for the upwards slope of the short-run AS curve, carefully explain the following.
 - a. How the economy recovers from a recession and returns to its long-run equilibrium without any policy intervention.
 - b. What determines the speed of that recovery?
- 7 Suppose the central bank expands the money supply, but because the public expects this action, it simultaneously raises its expectation of the price level. What will happen to output and the price level in the short run? Compare this result to the outcome if the central bank expanded the money supply but the public didn't change its expectation of the price level.
- 8 Suppose workers and firms suddenly believe that inflation will be quite high over the coming year. Suppose also that the economy begins in long-run equilibrium, and the AD curve does not shift.
 - a. What happens to nominal wages? What happens to real wages?
 - b. Using an AD/AS diagram, show the effect of the change in expectations on both the short-run and long-run levels of prices and output.
 - c. Were the expectations of high inflation accurate? Explain.
- 9 Explain whether each of the following events shifts the short-run AS curve, the AD curve, both, or neither. For each event that does shift a curve, use a diagram to illustrate the effect on the economy.
 - a. Households decide to save a larger share of their income.
 - b. Cattle farmers suffer a prolonged period of foot-and-mouth disease which cuts average cattle herd sizes by 80 per cent.
 - c. Increased job opportunities overseas cause many people to leave the country.
- 10 Suppose that firms become very optimistic about future business conditions and invest heavily in new capital equipment.
 - a. Use an AD/AS diagram to show the short-run effect of this optimism on the economy. Label the new levels of prices and real output. Explain, in words, why the aggregate quantity of output supplied changes.
 - b. Now use the diagram from part (a) to show the new long-run equilibrium of the economy. (For now, assume there is no change in the long-run AS curve.) Explain, in words, why the aggregate quantity of output demanded changes between the short run and the long run.
 - c. How might the investment boom affect the long-run AS curve? Explain.

33

THE INFLUENCE OF MONETARY AND FISCAL POLICY ON AGGREGATE DEMAND

Imagine that you are a member of a central bank's monetary policy committee (MPC), which sets a country's monetary policy. You observe that the Finance Minister has announced in his budget speech that he is going to cut government spending. How should the MPC respond to this change in fiscal policy? Should it reduce interest rates (and so allow the money supply to expand faster), raise interest rates or leave interest rates the same?

To answer this question, you need to consider the impact of monetary and fiscal policy on the economy. In the preceding chapter we saw how to explain short-run economic fluctuations using the model of aggregate demand and aggregate supply. When the aggregate demand curve or the aggregate supply curve shifts, the result is fluctuations in the economy's overall output of goods and services and in its overall level of prices. As we noted in the previous chapter, monetary and fiscal policy can each influence aggregate demand. Thus, a change in one of these policies can lead to short-run fluctuations in output and prices. Policymakers will want to anticipate this effect and, perhaps, adjust the other policy in response.

In this chapter we examine in more detail how the tools of monetary and fiscal policy influence the position of the aggregate demand curve. We have previously discussed the long-run effects of these policies. In Chapters 22 and 23 we saw how fiscal policy affects saving, investment and long-run economic growth. In Chapters 26 and 27 we saw how the central bank controls the money supply and how the money supply affects the price level in the long run. We now see how these policy tools can shift the aggregate demand curve and, in doing so, affect short-run economic fluctuations.

As we have already learned, many factors influence aggregate demand besides monetary and fiscal policy. In particular, desired spending by households and firms determines the overall demand for goods and services. When desired spending changes, aggregate demand shifts. If policymakers do not respond, such shifts in aggregate demand cause short-run fluctuations in output and employment. As a result, monetary and fiscal policymakers sometimes use the policy levers at their disposal to try to offset these shifts in aggregate demand and thereby stabilize the economy. Here we discuss the theory behind these policy actions and some of the difficulties that arise in using this theory in practice.

HOW MONETARY POLICY INFLUENCES AGGREGATE DEMAND

The aggregate demand curve shows the total quantity of goods and services demanded in the economy for any price level. As you may recall from the preceding chapter, the aggregate demand curve slopes downwards for three reasons:

- *The wealth effect.* A lower price level raises the real value of households' money holdings, and higher real wealth stimulates consumer spending.
- *The interest rate effect.* A lower price level lowers the interest rate as people try to lend out their excess money holdings, and the lower interest rate stimulates investment spending.
- *The exchange rate effect.* When a lower price level lowers the interest rate, investors move some of their funds overseas and cause the domestic currency to depreciate relative to foreign currencies. This depreciation makes domestic goods cheaper compared to foreign goods and, therefore, stimulates spending on net exports.

These three effects should not be viewed as alternative theories. Instead, they occur simultaneously to increase the quantity of goods and services demanded when the price level falls and to decrease it when the price level rises.

Although all three effects work together in explaining the downwards slope of the aggregate demand curve, they are not of equal importance. Because money holdings are a small part of household wealth, the wealth effect is the least important of the three. Whether or not the exchange rate effect is important depends upon the degree of openness of the economy. For a relatively closed economy like the USA, for example, because exports and imports represent only a small fraction of US GDP, the exchange rate effect is not very large. In the UK and most other European economies, however, the economy is much more open in the sense that imports and exports represent a much larger fraction of total GDP, so that the exchange rate effect will be more important. However, the exchange rate effect is probably secondary to the interest rate effect even in relatively open economies, for two reasons. First, the interest rate effect impacts immediately upon the whole economy, affecting consumers, homebuyers and firms across the board, while the loss of competitiveness that is part of the exchange rate effect impacts only upon traded goods, and affects mainly firms producing tradable goods and consumers buying tradable goods. Secondly, many of the countries of Europe have a common currency – the euro – which they share with many of their major trading partners, so that the exchange rate effect is muted.

To understand how policy influences aggregate demand, therefore, we shall examine the interest rate effect in more detail. Here we develop a theory of how the interest rate is determined, called the **theory of liquidity preference**, which was originally developed by John Maynard Keynes in the 1930s. After we develop this theory, we use it to understand the downward slope of the aggregate demand curve and how monetary policy shifts this curve. By shedding new light on the aggregate demand curve, the theory of liquidity preference expands our understanding of short-run economic fluctuations.

theory of liquidity preference Keynes' theory that the interest rate adjusts to bring money supply and money demand into balance

The Theory of Liquidity Preference

In his classic book, *The General Theory of Employment, Interest and Money*, John Maynard Keynes proposed the theory of liquidity preference to explain what factors determine the economy's interest rate. The theory is, in essence, just an application of supply and demand. According to Keynes, the interest rate adjusts to balance the supply and demand for money. We introduced some elements of the theories covered here in Chapter 31 as part of the IS-LM model.

You may recall that economists distinguish between two interest rates: the *nominal interest rate* is the interest rate as usually reported, and the *real interest rate* is the interest rate corrected for the effects of inflation. Which interest rate are we now trying to explain? The answer is both. In the analysis that follows, we hold constant the expected rate of inflation. (This assumption is reasonable for studying the economy in the short run, as we are now doing.) Thus, when the nominal interest rate rises or falls, the real interest rate that people expect to earn rises or falls as well. For the rest of this chapter, when we refer to changes in the interest rate, you should envision the real and nominal interest rates moving in the same direction.

Let's now develop the theory of liquidity preference by considering the supply and demand for money and how each depends on the interest rate.

Money Supply The first element of the theory of liquidity preference is the supply of money. As we have previously noted, the money supply is controlled by the central bank, such as the Bank of England, the European Central Bank or the US Federal Reserve. The central bank can alter the money supply by changing the quantity of reserves in the banking system through the purchase and sale of government bonds in outright open-market operations. When the central bank buys government bonds, the money it pays for the bonds is typically deposited in banks, and this money is added to bank reserves. When the central bank sells government bonds, the money it receives for the bonds is withdrawn from the banking system,

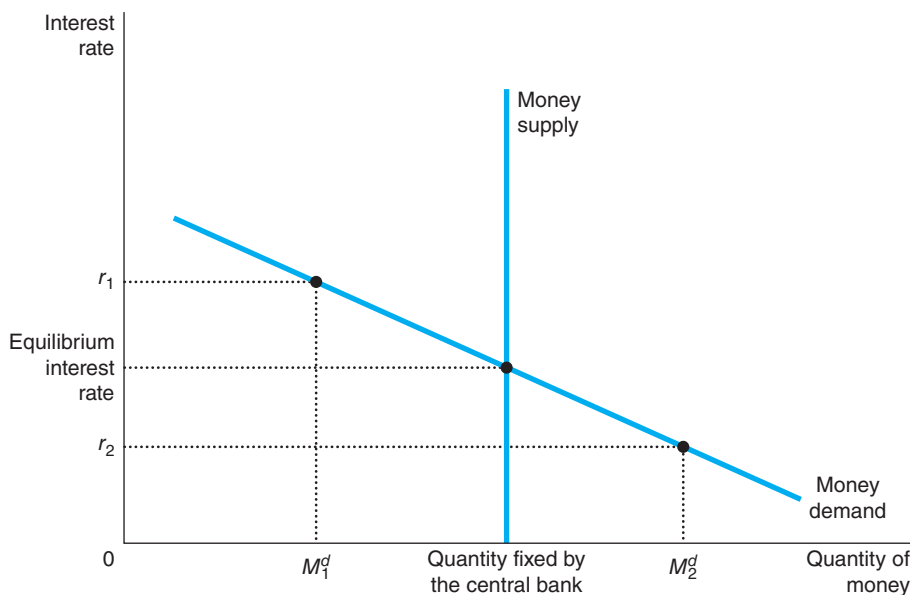
and bank reserves fall. These changes in bank reserves, in turn, lead to changes in banks' ability to make loans and create money. In addition to these open-market operations, the central bank can alter the money supply by changing reserve requirements (the amount of reserves banks must hold against deposits) or the refinancing rate (the interest rate at which banks can borrow reserves from the central bank).

These details of monetary control are important for the implementation of central bank policy, and we discussed them in detail in Chapter 26, but they are not crucial in this chapter. Our goal here is to examine how changes in the money supply affect the aggregate demand for goods and services. For this purpose, we can ignore the details of how central bank policy is implemented and simply assume that the central bank controls the money supply directly. In other words, the quantity of money supplied in the economy is fixed at whatever level the central bank decides to set it.

Because the quantity of money supplied is fixed by central bank policy, it does not depend on other economic variables. In particular, it does not depend on the interest rate. Once the central bank has made its policy decision, the quantity of money supplied is the same, regardless of the prevailing interest rate. We represent a fixed money supply with a vertical supply curve in Figure 33.1.

FIGURE 33.1**Equilibrium in the Money Market**

According to the theory of liquidity preference, the interest rate adjusts to bring the quantity of money supplied and the quantity of money demanded into balance. If the interest rate is above the equilibrium level (such as at r_1), the quantity of money people want to hold (M_1^d) is less than the quantity the central bank has created, and this surplus of money puts downward pressure on the interest rate. Conversely, if the interest rate is below the equilibrium level (such as at r_2), the quantity of money people want to hold (M_2^d) is greater than the quantity the central bank has created, and this shortage of money puts upward pressure on the interest rate. Thus, the forces of supply and demand in the market for money push the interest rate towards the equilibrium interest rate, at which people are content holding the quantity of money the central bank has created.



Money Demand The second element of the theory of liquidity preference is the demand for money. As a starting point for understanding money demand, recall that any asset's *liquidity* refers to the ease with which that asset is converted into the economy's medium of exchange. Money is the economy's medium of exchange, so it is by definition the most liquid asset available. The liquidity of money explains the demand for it: people choose to hold money instead of other assets that offer higher rates of return because money can be used to buy goods and services.

Although many factors determine the quantity of money demanded, the one emphasized by the theory of liquidity preference is the interest rate. The reason is that the interest rate is the opportunity cost of holding money. That is, when you hold wealth as cash in your pocket, instead of as an interest-bearing bond or bank account, you forgo the benefits of the interest you could have earned (the opportunity cost). An increase in the interest rate raises the opportunity cost of holding money. There is an incentive, therefore, for people to exchange cash holdings for interest-bearing deposits and this, as a result, reduces the quantity of money demanded. A decrease in the interest rate reduces the opportunity cost of holding money. The cost of the benefits forgone are not as high so there is more incentive to hold money as cash and as a result the quantity demanded increases. Thus, as shown in Figure 33.1, the money demand curve slopes downwards.

Equilibrium in the Money Market According to the theory of liquidity preference, the interest rate adjusts to balance the supply and demand for money. There is one interest rate, called the *equilibrium interest rate*, at which the quantity of money demanded exactly balances the quantity of money supplied. If the interest rate is at any other level, people will try to adjust their portfolios of assets and, as a result, drive the interest rate towards the equilibrium.

For example, suppose that the interest rate is above the equilibrium level, such as r_1 in Figure 33.1. In this case, the quantity of money that people want to hold, (M_1^d), is less than the quantity of money that the central bank has supplied. Those people who are holding the surplus of money will try to get rid of it by buying interest-bearing bonds or by depositing it in an interest-bearing bank account. Because bond issuers and banks prefer to pay lower interest rates, they respond to this surplus of money by lowering the interest rates they offer. As the interest rate falls, people become more willing to hold money until, at the equilibrium interest rate, people are happy to hold exactly the amount of money the central bank has supplied.

Conversely, at interest rates below the equilibrium level, such as r_2 in Figure 33.1, the quantity of money that people want to hold, (M_2^d), is greater than the quantity of money that the central bank has supplied. As a result, people try to increase their holdings of money by reducing their holdings of bonds and other interest-bearing assets. As people cut back on their holdings of bonds, bond issuers find that they have to offer higher interest rates to attract buyers. Thus, the interest rate rises and approaches the equilibrium level.

The Downwards Slope of the Aggregate Demand Curve

Having seen how the theory of liquidity preference explains the economy's equilibrium interest rate, we now consider its implications for the aggregate demand for goods and services. As a warm-up exercise, let's begin by using the theory to re-examine a topic we already understand – the interest rate effect and the downwards slope of the aggregate demand curve. In particular, suppose that the overall level of prices in the economy rises. What happens to the interest rate that balances the supply and demand for money, and how does that change affect the quantity of goods and services demanded?

As we discussed in Chapter 27, the price level is one determinant of the quantity of money demanded. At higher prices, more money is exchanged every time a good or service is sold. As a result, people will choose to hold a larger quantity of money. That is, a higher price level increases the quantity of money demanded for any given interest rate. Thus, an increase in the price level from P_1 to P_2 shifts the money demand curve to the right from MD_1 to MD_2 , as shown in panel (a) of Figure 33.2.

Notice how this shift in money demand affects the equilibrium in the money market. For a fixed money supply, the interest rate must rise to balance money supply and money demand. The higher price level has increased the amount of money people want to hold and has shifted the money demand curve to the right. Yet the quantity of money supplied is unchanged, so the interest rate must rise from r_1 to r_2 to discourage the additional demand.

This increase in the interest rate has ramifications not only for the money market but also for the quantity of goods and services demanded, as shown in panel (b). At a higher interest rate, the cost of borrowing and the return to saving are greater. Fewer households choose to borrow to buy a

FYI



Interest Rates in the Long Run and the Short Run

At this point, we should pause and reflect on a seemingly awkward embarrassment of riches. It might appear as if we now have two theories for how interest rates are determined. Chapter 23 states that the interest rate adjusts to balance the supply of and demand for loanable funds (that is, national saving and desired investment). By contrast, we just established here that the interest rate adjusts to balance the supply and demand for money. How can we reconcile these two theories?

To answer this question we must again consider the differences between the long-run and short-run behaviour of the economy. Three macroeconomic variables are of central importance: the economy's output of goods and services; the interest rate; and the price level. According to the classical macroeconomic theory we developed in Chapters 22, 23 and 24, these variables are determined as follows:

1. *Output* is determined by the supplies of capital and labour and the available production technology for turning capital and labour into output. (We call this the natural rate of output.)
2. For any given level of output, the *interest* rate adjusts to

balance the supply and demand for loanable funds.

3. The *price* level adjusts to balance the supply and demand for money. Changes in the supply of money lead to proportionate changes in the price level.

These are three of the essential propositions of classical economic theory. Most economists believe that these propositions do a good job of describing how the economy works *in the long run*.

Yet these propositions do not hold in the short run. As we discussed in the preceding chapter, many prices are slow to adjust to changes in the money supply; this is reflected in a short-run aggregate supply curve that is upwards sloping rather than vertical. As a result, the overall price level cannot, by itself, balance the supply and demand for money in the short run. This stickiness of the price level forces the interest rate to move in order to bring the money market into equilibrium. These changes in the interest rate, in turn, affect the aggregate demand for goods and services. As aggregate demand fluctuates, the economy's output of goods and services moves away from the level determined by factor supplies and technology.

For issues concerning the short run, then, it is best to think about the economy as follows:

1. The *price level* is stuck at some level (based on previously formed expectations) and, in the short run, is relatively unresponsive to changing economic conditions.
2. For any given price level, the *interest rate* adjusts to balance the supply and demand for money.
3. The level of *output* responds to the aggregate demand for goods and services, which is in part determined by the interest rate that balances the money market.

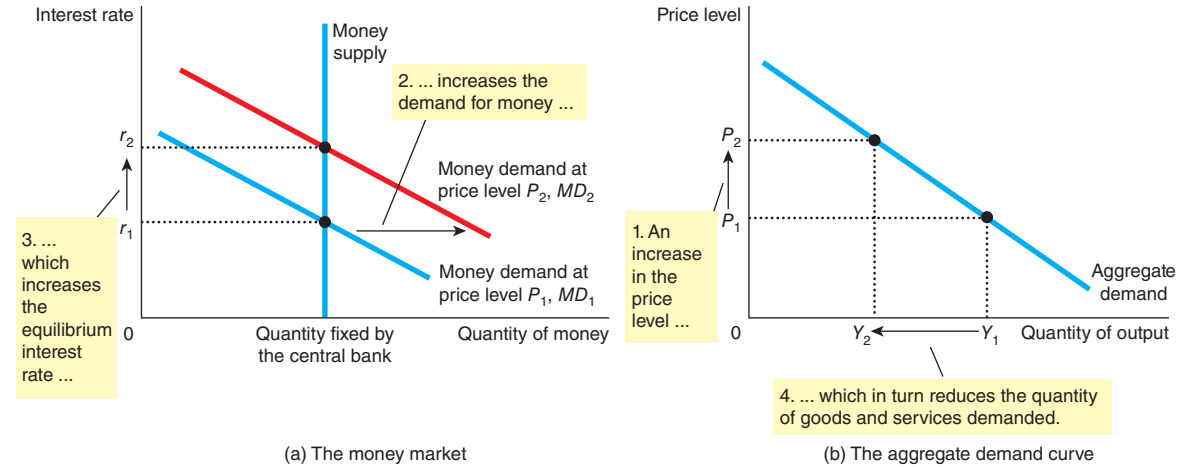
Notice that this precisely reverses the order of analysis used to study the economy in the long run.

Thus, the different theories of the interest rate are useful for different purposes. When thinking about the long-run determinants of interest rates, it is best to keep in mind the loanable funds theory. This approach highlights the importance of an economy's saving propensities and investment opportunities. By contrast, when thinking about the short-run determinants of interest rates, it is best to keep in mind the liquidity preference theory. This theory highlights the importance of monetary policy.

new house, and those who do buy smaller houses, so the demand for residential investment falls. Fewer firms choose to borrow to build new factories and buy new equipment, so business investment falls. Thus, when the price level rises from P_1 to P_2 , increasing money demand from MD_1 to MD_2 and raising the interest rate from r_1 to r_2 , the quantity of goods and services demanded falls from Y_1 to Y_2 .

FIGURE 33.2**The Money Market and the Slope of the Aggregate Demand Curve**

An increase in the price level from P_1 to P_2 shifts the money demand curve to the right, as in panel (a). This increase in money demand causes the interest rate to rise from r_1 to r_2 . Because the interest rate is the cost of borrowing, the increase in the interest rate reduces the quantity of goods and services demanded from Y_1 to Y_2 . This negative relationship between the price level and quantity demanded is represented with a downwards sloping aggregate demand curve, as in panel (b).



Hence, this analysis of the interest rate effect can be summarized in three steps:

1. A higher price level raises money demand.
2. Higher money demand leads to a higher interest rate.
3. A higher interest rate reduces the quantity of goods and services demanded.

Of course, the same logic works in reverse as well: a lower price level reduces money demand, which leads to a lower interest rate, and this in turn increases the quantity of goods and services demanded. The end result of this analysis is a negative relationship between the price level and the quantity of goods and services demanded, which is illustrated with a downwards sloping aggregate demand curve. This is the same conclusion we reached when looking at the effect of a shift in the LM curve in Chapter 31.

Changes in the Money Supply

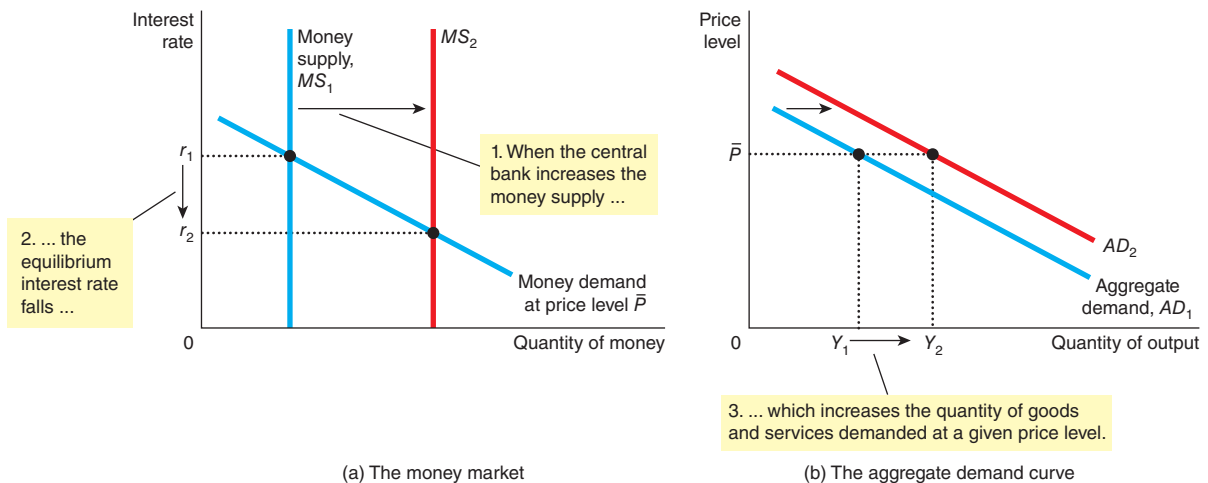
So far we have used the theory of liquidity preference to explain more fully how the total quantity demanded of goods and services in the economy changes as the price level changes. That is, we have examined movements along the downwards sloping aggregate demand curve. The theory also sheds light, however, on some of the other events that alter the quantity of goods and services demanded. Whenever the quantity of goods and services demanded changes *for a given price level*, the aggregate demand curve shifts.

One important variable that shifts the aggregate demand curve is monetary policy. To see how monetary policy affects the economy in the short run, suppose that the central bank increases the money supply by buying government bonds in open-market operations. (Why the central bank might do this will become clear later after we understand the effects of such a move.) Let's consider how this monetary injection influences the equilibrium interest rate for a given price level. This will tell us what the injection does to the position of the aggregate demand curve.

As panel (a) of Figure 33.3 shows, an increase in the money supply shifts the money-supply curve to the right from MS_1 to MS_2 . Because the money demand curve has not changed, the interest rate falls from r_1 to r_2 to balance money supply and money demand. That is, the interest rate must fall to induce people to hold the additional money that the central bank has created.

FIGURE 33.3**A Monetary Injection**

In panel (a), an increase in the money supply from MS_1 to MS_2 reduces the equilibrium interest rate from r_1 to r_2 . Because the interest rate is the cost of borrowing, the fall in the interest rate raises the quantity of goods and services demanded at a given price level from Y_1 to Y_2 . Thus, in panel (b), the aggregate demand curve shifts to the right from AD_1 to AD_2 .



Once again, the interest rate influences the quantity of goods and services demanded, as shown in panel (b) of Figure 33.3. The lower interest rate reduces the cost of borrowing and the return to saving. Households buy more and larger houses, stimulating the demand for residential investment. Firms spend more on new factories and new equipment, stimulating business investment. As a result, the quantity of goods and services demanded at a given price level \bar{P} rises from Y_1 to Y_2 . Of course, there is nothing special about \bar{P} : the monetary injection raises the quantity of goods and services demanded at every price level. Thus, the entire aggregate demand curve shifts to the right.

To sum up: when the central bank increases the money supply, it leads to a fall in the interest rate and increases the quantity of goods and services demanded for any given price level, shifting the aggregate demand curve to the right. Conversely, when the central bank contracts the money supply, the interest rate rises to bring the money market into equilibrium and reduces the quantity of goods and services demanded for any given price level, shifting the aggregate demand curve to the left.

We explored the same principle in Chapter 31 as part of the introduction to the IS-LM model. In that case a loosening of monetary policy causes the LM curve to shift to the right and at the given price level the aggregate demand curve shifts to the right (and vice versa).

The Role of Interest Rates

Our discussion so far in this chapter has treated the money supply as the central bank's policy instrument. When the central bank buys government bonds in open-market operations, it increases the money supply and expands aggregate demand. When the central bank sells government bonds in open-market operations, it decreases the money supply and contracts aggregate demand.

Often, however, discussions of central bank policy treat the interest rate, rather than the money supply, as the central bank's policy instrument. Indeed, in recent years many of the major central banks, including the Bank of England, the European Central Bank and the US Federal Reserve, have conducted policy by setting the interest rate at which they will lend to the banking sector – the refinancing rate for the European Central Bank, the repurchase or 'repo' rate for the Bank of England, and the discount rate for the Federal Reserve.

The central bank's decision to set interest rates rather than target a certain level (or rate of growth) of the money supply does not fundamentally alter our analysis of monetary policy. The theory of liquidity preference illustrates an important principle: monetary policy can be described either in terms of the money supply or in terms of the interest rate. When the central bank sets a target for the refinancing rate of, say, x per cent, the central bank's bond traders are told: 'Conduct whatever open-market operations are necessary to ensure that the equilibrium interest rate equals x per cent.' In other words, when the central bank sets a target for the interest rate, it commits itself to adjusting the money supply in order to make the equilibrium in the money market hit that target.

As a result, changes in monetary policy can be viewed either in terms of a changing target for the interest rate or in terms of a change in the money supply. As noted in Chapter 26, when you read in the newspaper that the central bank has lowered interest rates, you should understand that this occurs only because the central bank's bond traders are doing what it takes to make it happen. If interest rates have been lowered, then the central bank's bond traders will have bought government bonds, and this purchase increases the money supply and lowers the equilibrium interest rate (just as in Figure 33.3).

The lessons from all this are quite simple: changes in monetary policy that aim to expand aggregate demand can be described either as increasing the money supply or as lowering the interest rate. Changes in monetary policy that aim to contract aggregate demand can be described either as decreasing the money supply or as raising the interest rate.

SELF TEST Use the theory of liquidity preference to explain how a decrease in the money supply affects the equilibrium interest rate. How does this change in monetary policy affect the demand curve?

HOW FISCAL POLICY INFLUENCES AGGREGATE DEMAND

The government can influence the behaviour of the economy not only with monetary policy but also with fiscal policy. Fiscal policy refers to the government's choices regarding the overall level of government purchases or taxes. Earlier in the book we examined how fiscal policy influences saving, investment and growth in the long run. In the short run, however, the primary effect of fiscal policy is on the aggregate demand for goods and services.

Changes in Government Purchases

We saw in Chapter 31 that changes in autonomous spending can have an effect on the level of spending in the economy which is greater than the initial injection. The multiplier effect means that aggregate demand will shift by a larger amount than the increase in government spending. However, the crowding-out effect, which we introduced in Chapter 24, suggests that the shift in aggregate demand could be *smaller* than the initial injection.

The Crowding-Out Effect

While an increase in government purchases stimulates the aggregate demand for goods and services, it also causes the interest rate to rise, and a higher interest rate reduces investment spending and chokes off aggregate demand. The reduction in aggregate demand that results when a fiscal expansion raises the interest rate is called the **crowding-out effect**.

crowding-out effect the offset in aggregate demand that results when expansionary fiscal policy raises the interest rate and thereby reduces investment spending

CASE STUDY Negative Interest Rates and 'Crank' Ideas

Look around most developed countries in early 2013 and central bank interest rates are at historically low levels and have been so for some years following the financial crisis. Some central banks have also tried quantitative easing and other methods to try and get cash into the economy and trigger consumption and investment. There is some evidence to suggest that in many countries monetary policy is proving impotent in countering the fiscal austerity programmes governments have adopted and as a result economic growth remains sluggish. The Bank of England's QE programme, it is argued, has simply led to banks building up their balance sheets but not extending their lending to businesses and consumers to the extent that the Bank of England and the government would like.

If the traditional levers of monetary policy are not working might more creative ways to achieve desired policy objectives be used instead? In 2009, Professor Mankiw wrote an article in the *New York Times* suggesting negative interest rates. Borrow €1,000 today and pay back a smaller amount at some point in the future. If the interest rate was -4 per cent and you had borrowed the €1,000 for a period of a year, you would pay back €960 in a year's time. In that same article, Professor Mankiw recounted a discussion with one of his graduate students at Harvard about a scheme, put forward by the student, whereby the central bank announces that in one year's time it would pick a digit from one to nine out of a hat and any currency ending in that number would cease to be legal tender. People would thus know that in one year's time 10 per cent of the cash would cease to be legal tender, so what would they do? The logic is to spend it. The additional spending would increase aggregate demand and act as a boost to the economy. Such a policy might enable central banks to set negative interest rates provided the rate was less than 10 per cent because there would then be an incentive to lend at (say) -4 per cent rather than potentially losing 10 per cent.

Mankiw noted that the basic idea of a negative interest rate was not new – a late 19th-century economist, Silvio Gesell, had mooted such an idea and Gesell's idea was picked up by Keynes. Keynes noted that initially he saw Gesell's ideas as being those of a 'crank' but changed his opinion and even referred to Gesell as an 'unduly neglected prophet'.



In 2013, the idea of negative interest rates again surfaced when the Bank of England deputy governor outlined the possibility, albeit in very cautious terms, when giving evidence to the Treasury Select Committee. If the Bank of England reduced its lending rate to the banking system to negative rates then banks would be penalized for holding cash and would be incentivized to lend it out to avoid the penalty.

Economists expressed some surprise at the suggestion by Mr Tucker and many noted that it may have considerable unintended consequences and indeed Mr Tucker himself noted that it 'would be an extraordinary thing to do and it needs to be thought through very carefully'.

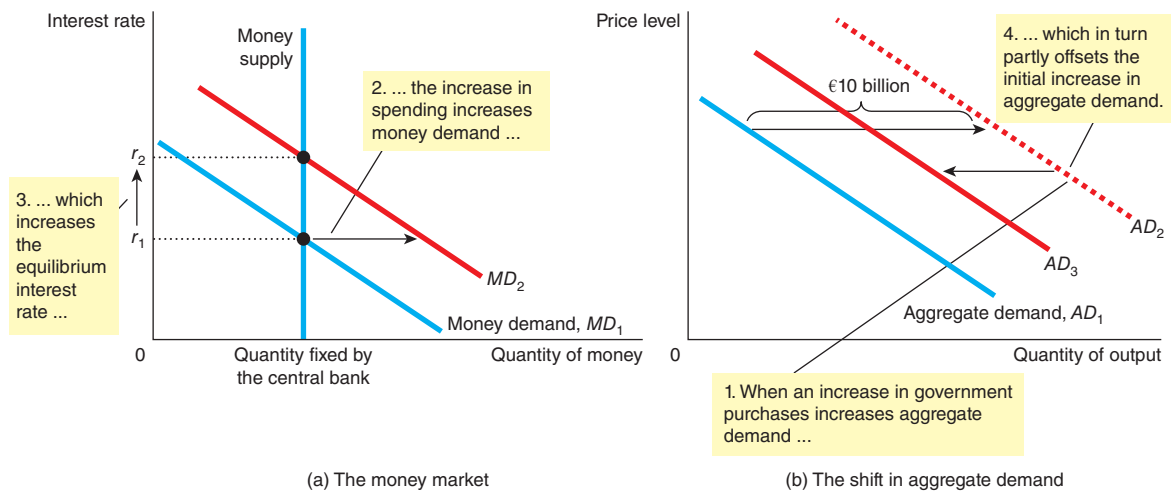
Former Bank of England deputy governor, Paul Tucker, mooted the idea of negative interest rates and whilst it has some appeal, there may be some unintended consequences as a result of one of the Ten Principles of Economics – people respond to incentives.

To see why the crowding-out effect occurs, let's consider what happens in the money market using an example we used in Chapter 31 when the government invests in nuclear power stations from Nucelec. As we discussed in Chapter 31, this increase in demand raises the incomes of the workers and owners of this firm (and, because of the multiplier effect, of other firms as well). As incomes rise, households plan to buy more goods and services and, as a result, choose to hold more of their wealth in liquid form. That is, the increase in income caused by the fiscal expansion raises the demand for money.

The effect of the increase in money demand is shown in panel (a) of Figure 33.4. Because the central bank has not changed the money supply, the vertical supply curve remains the same. When the higher level of income shifts the money demand curve to the right from MD_1 to MD_2 , the interest rate must rise from r_1 to r_2 to keep supply and demand in balance.

FIGURE 33.4**The Crowding-Out Effect**

Panel (a) shows the money market. When the government increases its purchases of goods and services, the resulting increase in income raises the demand for money from MD_1 to MD_2 , and this causes the equilibrium interest rate to rise from r_1 to r_2 . Panel (b) shows the effects on aggregate demand. The initial impact of the increase in government purchases shifts the aggregate demand curve from AD_1 to AD_2 . Yet, because the interest rate is the cost of borrowing, the increase in the interest rate tends to reduce the quantity of goods and services demanded, particularly for investment goods. This crowding out of investment partially offsets the impact of the fiscal expansion on aggregate demand. In the end, the aggregate demand curve shifts only to AD_3 .



The increase in the interest rate, in turn, reduces the quantity of goods and services demanded. In particular, because borrowing is more expensive, the demand for residential and business investment goods declines. That is, as the increase in government purchases increases the demand for goods and services, it may also crowd out investment. This crowding-out effect partially offsets the impact of government purchases on aggregate demand, as illustrated in panel (b) of Figure 33.4. The initial impact of the increase in government purchases is to shift the aggregate demand curve from AD_1 to AD_2 , but once crowding out takes place, the aggregate demand curve drops back to AD_3 .

To sum up: when the government increases its purchases by €10 billion, the aggregate demand for goods and services could rise by more or less than €10 billion, depending on whether the multiplier effect or the crowding-out effect is larger.

Changes in Taxes

The other important instrument of fiscal policy, besides the level of government purchases, is the level of taxation. When the government cuts personal income taxes, for instance, it increases households' take-home pay. Households will save some of this additional income, but they will also spend some of it on consumer goods. Because it increases consumer spending, the tax cut shifts the aggregate demand curve to the right. Similarly, a tax increase depresses consumer spending and shifts the aggregate demand curve to the left.

FYI



The Taylor Rule

In 1993 John Taylor, an economist at Stanford University in the United States, spent some time observing the behaviour of the Federal Reserve's Federal Open Market Committee (FOMC). Taylor's observations led him to put forward what has become known as the *Taylor rule* which has been an influential idea since it was published. For many economists, interest has centred on the extent to which the rule is followed in interest rate decision-making around the world, if at all. John Taylor was observing the US economy and the work of the FOMC. Inflation in the US had come under a far greater degree of control in the 1980s when the FOMC was headed by Paul Volcker. Part of Taylor's conclusion was that the FOMC had reacted more aggressively to inflation than had occurred before 1979. His observations led to the notion of the interest rate-setting body 'leaning into the wind' when it comes to inflation. By this he meant being willing to raise rates more significantly in response to an inflation threat than had been the case prior to 1979.

Taylor suggested that the pattern of FOMC behaviour with regard to interest rates during the period 1979–1992 could be expressed as a formula – the 'rule'. This formula, given below, has been simplified but in essence captures the main flavour of his original work.

$$r = p + 0.5y + 0.5(p - 2) + 2$$

In this formula, the following variables are expressed:

r = the short-term interest rate in percentage terms per annum

p = the rate of inflation over the previous four quarters measured as the GDP deflator

y = the difference between real GDP from potential output (the 'output gap').

Taylor made a number of assumptions about the US economy in this formula. He assumed that there was a target level of inflation, which he put at 2 per cent. He also made an assumption that the equilibrium real interest rate was also 2 per cent. The formula offers a rule or guide to policymakers about what the level of interest rates should be if there is a target for inflation of 2 per cent. In effect, interest rates can be set in response to the deviation of inflation from the target and the deviation of real output from potential output. It suggests that the response to inflation or output being off-target should be met with a more aggressive response to monetary policy – this is the so-called 'leaning into the wind'.

Suppose, for example, that the output gap was zero – that actual GDP was at its potential level – but that inflation was currently 5 per cent, 3 per cent above the assumed target level of 2 per cent. Then the *Taylor rule* would suggest that the central bank should aim to set the short-term interest rate at 8.5 per cent, corresponding to a real rate of

interest of $8.5 - 5 = 3.5$ per cent. If, however, GDP was, say, 2 per cent below its potential level, so that there was an output gap of -2 per cent, then the *Taylor rule* would suggest setting the nominal interest rate a little lower, at 7.5 per cent.

How accurate is the *Taylor rule*? Studies have shown that a number of countries seem to have patterns of interest rates that follow the *Taylor rule* relatively closely. In reality, Taylor and other researchers appreciated that decision makers could not stick rigidly to such a rule because sometimes circumstances might dictate the need to apply discretion. Such circumstances might be, for example, when the World Trade Center was attacked in 2001, the intervention in Iraq or the financial crisis of 2007–2009. Taylor argued that whilst a rule gave a guideline to policymakers, it also implied that any deviation from the rule had to be explained through a coherent and well-argued case and that such a discipline was helpful to overall monetary policy decisions. The *Taylor rule* has provided researchers with a great deal of impetus to look at the whole way in which monetary policy decisions are arrived at. It also gives the financial markets the opportunity to be able to second-guess the decision making of central banks and factor this into their thinking. In times of financial crisis, however, the rule may be jettisoned in favour of a more pragmatic approach to policy.

The size of the shift in aggregate demand resulting from a tax change is also affected by the multiplier and crowding-out effects. When the government cuts taxes and stimulates consumer spending, earnings and profits rise, which further stimulates consumer spending. This is the multiplier effect. At the same time, higher income leads to higher money demand, which tends to raise interest rates. Higher interest rates make borrowing more costly, which reduces investment spending. This is the crowding-out effect. Depending on the size of the multiplier and crowding-out effects, the shift in aggregate demand could be larger or smaller than the tax change that causes it.

In addition to the multiplier and crowding-out effects, there is another important determinant of the size of the shift in aggregate demand that results from a tax change: households' perceptions about whether the tax change is permanent or temporary. For example, suppose that the government announces a tax cut of €1,000 per household. In deciding how much of this €1,000 to spend, households must ask themselves how long this extra income will last. If households expect the tax cut to be permanent, they will view it as adding substantially to their financial resources and, therefore, increase their spending by a large amount. In this case, the tax cut will have a large impact on aggregate demand. By contrast, if households expect the tax change to be temporary, they will view it as adding only slightly to their financial resources and, therefore, will increase their spending by only a small amount. In this case, the tax cut will have a small impact on aggregate demand.

It should be noted that fiscal policy can also affect the aggregate supply curve and we will look at this in more detail in Chapter 35.

SELF TEST Suppose that the government reduces spending on motorway construction by €1 billion. Which way does the aggregate demand curve shift? Explain why the shift might be larger than €1 billion. Explain why the shift might be smaller than €1 billion.

USING POLICY TO STABILIZE THE ECONOMY

We have seen how monetary and fiscal policy can affect the economy's aggregate demand for goods and services. These theoretical insights raise some important policy questions: should policymakers use these instruments to control aggregate demand and stabilize the economy? If so, when? If not, why not?

The Case for an Active Stabilization Policy

Let's return to the question that began this chapter: when the government reduces its spending, how should the central bank respond? As we have seen, government spending is one determinant of the position of the aggregate demand curve. When the government cuts spending, aggregate demand will fall, this will depress production and employment in the short run. If the central bank wants to prevent this adverse effect of the fiscal policy, it can act to expand aggregate demand by increasing the money supply. A monetary expansion would reduce interest rates, stimulate investment spending and expand aggregate demand. If monetary policy responds appropriately, the combined changes in monetary and fiscal policy could leave the aggregate demand for goods and services unaffected.

This analysis is exactly the sort followed by the members of the policy-setting committees of central banks like the Bank of England, the European Central Bank and the Federal Reserve. They know that monetary policy is an important determinant of aggregate demand. They also know that there are other important determinants as well, including fiscal policy set by the government, and so they will watch debates over fiscal policy with a keen eye.

This response of monetary policy to the change in fiscal policy is an example of a more general phenomenon: the use of policy instruments to stabilize aggregate demand and, as a result, production and employment. We have seen in Chapter 31 how the Great Depression in the 1930s spawned a belief that governments should act to manage aggregate demand in the wake of the research which followed Keynes' *General Theory*. As a result, economic stabilization has been seen as an explicit or implicit goal of government macroeconomic policy in European and North American economies following World War II. In the UK, for example, this view was embodied in a government White Paper, published in 1944, which explicitly stated: 'The Government accepts as one of their primary aims and responsibilities the maintenance of a high and stable level of employment after the War.' In the USA, similar sentiments were embodied in the Employment Act of 1946. This explicit recognition by governments of a responsibility to stabilize the economy has two implications. The first, more modest, implication is that the government should avoid being a cause of economic fluctuations itself. Thus, most economists advise against large and sudden changes in monetary and fiscal policy, for such changes are likely to cause fluctuations in aggregate demand. Moreover, when large changes do occur, it is important that monetary and fiscal policymakers be aware of and respond to the other's actions.

The second, more ambitious, implication of this explicit admission of responsibility – and one that was especially dominant in the first 30 years after the end of World War II – was that the government should respond to changes in the private economy in order to stabilize aggregate demand. To put the reasons for the pursuit of active stabilization policies into context, it is important to note that politicians remembered the misery of the Great Depression before the war, and were keen to avoid a recurrence, not only because of the misery involved for millions of people, but also because of the political effects of economic depression which were associated with a rise in extremism. Poverty was directly linked with the rise of extremism and political instability. As the war ended, therefore, they wanted to look forward to a better world in which governments could help avoid major recessions and the associated political instability. The adoption of Keynesian demand management policies seemed to offer some promise that full employment could be achieved.

Keynes (and his many followers) argued that aggregate demand fluctuates because of largely irrational waves of pessimism and optimism. He used the term 'animal spirits' to refer to these arbitrary changes in attitude. When pessimism reigns, households reduce consumption spending, and firms reduce investment spending. The result is reduced aggregate demand, lower production and higher unemployment. Conversely, when optimism reigns, households and firms increase spending. The result is higher aggregate demand, higher production and inflationary pressure. Notice that these changes in attitude are, to some extent, self-fulfilling.

In principle, the government can adjust its monetary and fiscal policy in response to these waves of optimism and pessimism and, thereby, stabilize the economy. For example, when people are excessively pessimistic, the central bank can expand the money supply to lower interest rates and expand aggregate demand. When they are excessively optimistic, it can contract the money supply to raise interest rates and dampen aggregate demand.

The Case Against an Active Stabilization Policy

Some economists argue that the government should avoid active use of monetary and fiscal policy to try to stabilize the economy. They claim that these policy instruments should be set to achieve long-run goals, such as rapid economic growth and low inflation, and that the economy should be left to deal with short-run fluctuations on its own. Although these economists may admit that monetary and fiscal policy can stabilize the economy in theory, they doubt whether it can do so in practice.

The primary argument against active monetary and fiscal policy is that the effects of these policies may be to a large extent uncertain both in terms of magnitude and timing. As we have seen, monetary policy works by changing interest rates. This can have strong and rapid effects on consumer spending if (as in the United Kingdom) a large number of people are buying their house with a mortgage loan on which the interest rate can vary according to market interest rates: quite simply, if interest rates go up, then mortgage payments go up and people have less money to spend. Yet, if people have mortgages on which the interest rate is fixed for one or more years ahead, then the interest rate change will only affect mortgage

payments with a very long lag. And if most people live in rented accommodation (as is the case in most of the countries of continental Europe) then a rise in interest rates will have no strong effects through this channel at all. In all cases, however, the interest rate rise will still clearly affect consumer spending because buying goods on credit (e.g. with a credit card) will be more expensive. The net effect on consumer spending may, therefore, be hard to predict, especially in terms of its timing.

Monetary policy can also affect aggregate demand through its influence on investment spending. But many firms make investment plans far in advance. Thus, most economists believe that it takes at least six months for changes in monetary policy to have much effect on output and employment. Most central banks readily admit that changes in interest rates can take up to 18 months to work their way through the economy. Moreover, once these effects occur, they can last for several years.

Critics of stabilization policy argue that because of these uncertain lags, the central bank should not try to fine-tune the economy. They claim that the central bank often reacts too late to changing economic conditions and, as a result, ends up being a cause of, rather than a cure for, economic fluctuations. These critics advocate a passive monetary policy, such as slow and steady growth in the money supply.

Fiscal policy may also work with a lag. Of course, the impact of a change in government spending is felt as soon as the change takes place and cuts in direct and indirect taxation can feed through into the economy quickly. However, considerable time may pass between the decision to adopt a government spending programme and its implementation. In the UK, for example, the government has often tended to undershoot on its planned spending, partly because of problems in attracting sufficient extra staff into key public services such as transport, education and health. At the same time, announcements about changes to income taxes (both personal and corporate) tend to occur months (and sometimes years) before implementation. In the intervening time period, households and businesses factor in the impending changes into their behaviour and so the effects are uncertain.

These lags in monetary and fiscal policy are a problem in part because economic forecasting is so imprecise. If forecasters could accurately predict the condition of the economy a year in advance, then monetary and fiscal policymakers could look ahead when making policy decisions. In this case, policymakers could stabilize the economy despite the lags they face. Decisions are made, in part, on the basis of existing statistical data which are fed into models. That data can be inaccurate and subject to revision as we have seen. Without accurate and up-to-date information, the outcome from modelling can vary considerably. In practice, however, major recessions and depressions arrive without much advance warning. The best policymakers can do at any time is to respond to economic changes as they occur.

Automatic Stabilizers

All economists – both advocates and critics of stabilization policy – agree that the lags in implementation render policy less useful as a tool for short-run stabilization. The economy would be more stable, therefore, if policymakers could find a way to avoid some of these lags. In fact, they have. **Automatic stabilizers** are changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession without policymakers having to take any deliberate action.

automatic stabilizers changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession, without policymakers having to take any deliberate action

The most important automatic stabilizer is the tax system. When the economy goes into a recession, the amount of taxes collected by the government falls automatically because almost all taxes are closely tied to economic activity and because in many countries income taxes are progressive. This means that as economic activity increases an increasing proportion of income is paid in tax and vice versa. Income tax depends on households' incomes and corporation tax depends on firms' profits. Because incomes and profits both fall in a recession, the government's tax revenue falls as well. Taxes are a withdrawal

from the circular flow which has the effect of dampening the level of aggregate demand. If tax revenues are lower it means that consumers have more disposable income to spend on consumption and businesses on investment. If economic activity increases then tax revenues will rise. These automatic tax changes either stimulate or dampen aggregate demand and, thereby, reduce the magnitude of economic fluctuations.

Government spending also acts as an automatic stabilizer. In particular, when the economy goes into a recession and workers are laid off, more people apply for state unemployment benefits, welfare benefits and other forms of income support. Extra spending on benefits and the welfare system helps to provide a cushion against too large a fall in economic activity. This increase in government spending stimulates aggregate demand at exactly the time when aggregate demand is insufficient to maintain full employment. However, it must be taken into consideration that in order to finance this additional spending (at a time when tax revenue is falling) governments may have to borrow. This additional borrowing can put upward pressure on interest rates and dampen the overall effect.

Automatic stabilizers are generally not sufficiently strong to prevent recessions completely. Nevertheless, without these automatic stabilizers, output and employment would probably be more volatile than they are. For this reason, most economists would not favour a policy of always running a balanced budget, as some politicians have proposed. When the economy goes into a recession, taxes fall, government spending rises, and the government's budget moves toward deficit. If the government faced a strict balanced-budget rule, it would be forced to look for ways to raise taxes or cut spending in a recession. In other words, a strict **balanced budget** rule would eliminate the automatic stabilizers inherent in our current system of taxes and government spending and would, in effect, be an 'automatic destabilizer'.

balanced budget where the total sum of money received by a government in tax revenue and interest is equal to the amount it spends, including on any debt interest owing

SELF TEST Suppose a wave of negative 'animal spirits' overruns the economy, and people become pessimistic about the future. What happens to aggregate demand? If the central bank wants to stabilize aggregate demand, how should it alter the money supply? If it does this, what happens to the interest rate? Why might the central bank choose not to respond in this way?

CONCLUSION

Before policymakers make any change in policy, they need to consider all the effects of their decisions. Earlier in the book we examined classical models of the economy, which describe the long-run effects of monetary and fiscal policy. There we saw how fiscal policy influences saving, investment and long-run growth, and how monetary policy influences the price level and the inflation rate.

In this chapter we examined the short-run effects of monetary and fiscal policy. We saw how these policy instruments can change the aggregate demand for goods and services and, thereby, alter the economy's production and employment in the short run. When the government reduces spending in order to balance the budget, it needs to consider both the long-run effects on saving and growth and the short-run effects on aggregate demand and employment. When the central bank reduces the growth rate of the money supply, it must take into account the long-run effect on inflation as well as the short-run effect on production. In the next chapter we discuss the transition between the short run and the long run more fully, and we see that policymakers often face a trade-off between long-run and short-run goals.

IN THE NEWS



The Multiplier and Fiscal Policy

In Chapter 31 and in this chapter we have spoken of fiscal policy being used as a tool to boost the economy. Part of the justification for using government spending and taxation in this way is the role of the multiplier.

Fiscal Multipliers

The theory seems intuitive enough – if government spends money on a major infrastructure project it is logical that the income received will be spent and feed through to other parts of the economy and as a result lead to a rise in national income that is a multiple of the initial injection. Most economists would not disagree with this base analysis but where they do disagree is on the size of the multiplier effect and therefore whether this is the best way to smooth out short-term fluctuations in economic activity. There has been a considerable amount of research into the size of fiscal or government expenditure multipliers – the change in output that arises as a result of a change in government expenditure. In part this research has arisen because of the stimulus packages that governments put in place, the fact that monetary policy options have been reduced and the continued sluggishness of economies across Europe and North America (perhaps to a lesser extent) has led to calls for further fiscal stimulus at a time when many countries are struggling to manage government borrowing.

Does the size of the fiscal multiplier matter? Well, take the

difference in the estimates of the size of the fiscal multiplier according to two leading macroeconomists. Robert Barro, a Harvard professor, wrote in 2009 that the size of the fiscal multiplier in peacetime was close to zero; Christine Romer, a professor of economics at the University of California, Berkeley and a former Chair of Economic Advisors to President Obama, suggested that the multiplier was nearer to 1.6. When the US government passed a stimulus package of \$787 billion (€604 billion) in 2009, the difference in the number of jobs created of these two estimates of the multiplier effects would be around 3.75 million according to three economists, Mendoza, Vegh and Ilzetzki, in an article published on Vox, a policy portal of the Centre for Economic Policy Research.

What the research does seem to show is that the size of the fiscal multiplier is dependent on a wide range of factors. The fiscal multiplier might be different in developed and less-developed countries, whether the government spending ‘shock’ is anticipated or unanticipated, how open the economy is, the exchange rate regime used by the country (whether exchange rates are fixed against other currencies, managed or allowed to float freely

in response to market conditions), whether the economy is experiencing a financial or debt crisis and the time preferences of the population of a country. Time preferences refer to how people react to the fiscal stimulus; if a government cut taxes or increased spending, what proportion of the increase in income would be spent and what proportion saved and to what extent would people build into their decision-making the expectation that a ‘windfall’ now will ‘inevitably’ lead to tax rises in the future. The idea that people will save most of any increase in government spending because they expect to have to pay higher taxes in the future is called *Ricardian Equivalence* and was initially developed by David Ricardo in the 19th century but has been refined somewhat by Robert Barro. The size of the fiscal multiplier will depend in part, therefore, on the extent to which consumers are Ricardian in their response.

A discussion paper written by Gilberto Marcheggiano and David Miles (Bank of England External MPC Discussion Paper no. 39, January 2013) noted that ‘Empirical studies ... face great challenges in measuring multipliers because of the difficulties of identifying exogenous fiscal shocks and controlling

(Continued)

for other factors that might affect output responses'. They further note that studies on the size of fiscal multipliers include estimates between 1.2 and 1.8 and in other cases, 0.8 to 1.5.

Large-scale investment in infrastructure projects like this construction of a high-speed rail line between Turin in Italy and Lyon in France helps to create jobs and generates a multiplier effect.



Questions

1 How do you think the size of the fiscal multiplier would vary when a government announced an increase in spending of €50 billion to be spent on

infrastructure projects compared to a cut in income taxes which would put an equivalent amount into people's pockets?

2 Why do you think that Barro and Romer have such a different view of the size of the fiscal multiplier?

3 Why might the size of the fiscal multiplier depend on whether it is anticipated or unanticipated?

4 The size of the fiscal multiplier will depend on Ricardian Equivalence. To what extent do you think people adjust their spending decisions in response to expectations of future tax changes?

5 If it is difficult to empirically arrive at a reliable and accurate measure of the size of fiscal multipliers, does this mean that policymakers should avoid resorting to the use of fiscal policy as a means of stimulating an economy? Justify your answer.

SUMMARY

- In developing a theory of short-run economic fluctuations, Keynes proposed the theory of liquidity preference to explain the determinants of the interest rate. According to this theory, the interest rate adjusts to balance the supply and demand for money.
- An increase in the price level raises money demand and increases the interest rate that brings the money market into equilibrium. Because the interest rate represents the cost of borrowing, a higher interest rate reduces investment and, thereby, the quantity of goods and services demanded. The downwards sloping aggregate demand curve expresses this negative relationship between the price level and the quantity demanded.
- Policymakers can influence aggregate demand with monetary policy. An increase in the money supply reduces the equilibrium interest rate for any given price level. Because a lower interest rate stimulates investment spending, the aggregate demand curve shifts to the right. Conversely, a decrease in the money supply raises the equilibrium interest rate for any given price level and shifts the aggregate demand curve to the left.
- Policymakers can also influence aggregate demand with fiscal policy. An increase in government purchases or a cut in taxes shifts the aggregate demand curve to the right. A decrease in government purchases or an increase in taxes shifts the aggregate demand curve to the left.
- When the government alters spending or taxes, the resulting shift in aggregate demand can be larger or smaller than the fiscal change. The multiplier effect tends to amplify the effects of fiscal policy on aggregate demand. The crowding-out effect tends to dampen the effects of fiscal policy on aggregate demand.
- Because monetary and fiscal policy can influence aggregate demand, the government sometimes uses these policy instruments in an attempt to stabilize the economy. Economists disagree about how active the government should be in this effort. According

to advocates of active stabilization policy, changes in attitudes by households and firms shift aggregate demand; if the government does not respond, the result is undesirable and unnecessary fluctuations in output

and employment. According to critics of an active stabilization policy, monetary and fiscal policy work with such long lags that attempts at stabilizing the economy often end up being destabilizing.

QUESTIONS FOR REVIEW

- 1 What are the three ways in which monetary policy affects aggregate demand?
- 2 What is the theory of liquidity preference?
- 3 How does the theory of liquidity preference help explain the downwards slope of the aggregate demand curve?
- 4 Use the theory of liquidity preference to explain how a decrease in the money supply affects the aggregate demand curve.
- 5 The government spends €500 million to buy police cars. Explain why aggregate demand might increase by more than €500 million. Explain why aggregate demand might increase by less than €500 million.
- 6 What is the crowding-out effect?
- 7 Suppose that survey measures of consumer confidence indicate a wave of pessimism is sweeping the country. If policymakers do nothing, what will happen to aggregate demand? What should the government do if it wants to stabilize aggregate demand? If the government does nothing, should the central bank stabilize aggregate demand? If so, how?
- 8 How might a government use taxation as a tool in expansionary fiscal policy?
- 9 Outline the arguments for and against an active stabilization policy.
- 10 Give an example of a government policy that acts as an automatic stabilizer. Explain why this policy has this effect.

PROBLEMS AND APPLICATIONS

- 1 Explain how each of the following developments would affect the supply of money, the demand for money and the interest rate. Illustrate your answers with diagrams.
 - a. The central bank's bond traders buy bonds in open market operations.
 - b. An increase in credit card availability reduces the cash people hold.
 - c. The central bank reduces banks' reserve requirements.
 - d. Households decide to hold more money to use for holiday shopping.
 - e. A wave of optimism boosts business investment and expands aggregate demand.
 - f. An increase in oil prices shifts the short-run aggregate supply curve to the left.
- 2 Suppose banks install automatic teller machines on every street corner and, by making cash readily available, reduce the amount of money people want to hold.
 - a. Assume the central bank does not change the money supply. According to the theory of liquidity preference, what happens to the interest rate? What happens to aggregate demand?
 - b. If the central bank wants to stabilize aggregate demand, how should it respond?
- 3 The economy is in a recession with high unemployment and low output.
 - a. Use a graph of aggregate demand and aggregate supply to illustrate the current situation. Be sure to include the aggregate demand curve, the short-run aggregate supply curve, and the long-run aggregate supply curve.
 - b. Identify an open-market operation that would restore the economy to its natural rate.
 - c. Use a graph of the money market to illustrate the effect of this open-market operation. Show the resulting change in the interest rate.
 - d. Use a graph similar to the one in part (a) to show the effect of the open-market operation on output and the price level. Explain in words why the policy has the effect that you have shown in the graph.
- 4 This chapter explains that expansionary monetary policy reduces the interest rate and thus stimulates demand for investment goods. Explain how such a policy also stimulates the demand for net exports.
- 5 Suppose government spending increases. Would the effect on aggregate demand be larger if the central bank took no action in response, or if the central bank were committed to maintaining a fixed interest rate? Explain.

- 6 In which of the following circumstances is expansionary fiscal policy more likely to lead to a short-run increase in investment? Explain.
- When the investment accelerator is large, or when it is small?
 - When the interest sensitivity of investment is large, or when it is small?
- 7 Assume the economy is in a recession. Explain how each of the following policies would affect consumption and investment. In each case, indicate any direct effects, any effects resulting from changes in total output, any effects resulting from changes in the interest rate and the overall effect. If there are conflicting effects making the answer ambiguous, say so.
- An increase in government spending.
 - A reduction in taxes.
 - An expansion of the money supply.
- 8 For various reasons, fiscal policy changes automatically when output and employment fluctuate.
- Explain why tax revenue changes when the economy goes into a recession.
 - Explain why government spending changes when the economy goes into a recession.
 - If the government was to operate under a strict balanced-budget rule, what would it have to do in a recession? Would this make the recession more or less severe?
- 9 Recently, some members of the legislature have proposed a law that would make price stability the sole goal of monetary policy. Suppose such a law were passed.
- How would the central bank respond to an event that contracted aggregate demand?
 - How would the central bank respond to an event that caused an adverse shift in short-run aggregate supply?
 - In each case, is there another monetary policy that would lead to greater stability in output?
- 10 Some economists have proposed that central banks should use the following rule for choosing their target interest rate (r):
- $$r = 2\% + p + \frac{1}{2} \frac{(y - y^*)}{y^*} + \frac{1}{2} (p - p^*),$$
- where p is the average of the inflation rate over the past year, y is real GDP as recently measured, y^* is an estimate of the natural rate of output and p^* is the central bank's target rate of inflation.
- Explain the logic that might lie behind this rule for setting interest rates. Would you support the use of this rule?
 - Some economists advocate such a rule for monetary policy but believe p and y should be the forecasts of future values of inflation and output. What are the advantages of using forecasts instead of actual values? What are the disadvantages?

34

THE SHORT-RUN TRADE-OFF BETWEEN INFLATION AND UNEMPLOYMENT

Two closely watched indicators of economic performance in any advanced economy are inflation and unemployment. When the government agency responsible for producing national statistics releases data on these variables, policymakers are eager to hear the news. Sometimes commentators add together the inflation rate and the unemployment rate to produce a *misery index*, which purports to measure the health of the economy.

How are these two measures of economic performance related to each other? Earlier in the book we discussed the long-run determinants of unemployment and the long-run determinants of inflation. We saw that the natural rate of unemployment depends on various features of the labour market, such as minimum wage laws, the market power of unions, the role of efficiency wages and the effectiveness of job search. By contrast, the inflation rate depends primarily on growth in the money supply, which a nation's central bank controls. In the long run, therefore, inflation and unemployment are largely unrelated problems.

In the short run, just the opposite is true. One of the *Ten Principles of Economics* is that society faces a short-run trade-off between inflation and unemployment. If monetary and fiscal policymakers expand aggregate demand and move the economy up along the short-run aggregate supply curve, they can lower unemployment for a while, but only at the cost of higher inflation. If policymakers contract aggregate demand and move the economy down the short-run aggregate supply curve, they can lower inflation, but only at the cost of temporarily higher unemployment.

In this chapter we examine this trade-off more closely. The relationship between inflation and unemployment is a topic that has attracted the attention of some of the most important economists of the last half century. The best way to understand this relationship is to see how thinking about it has evolved over time.

THE PHILLIPS CURVE

'Probably the single most important macroeconomic relationship is the Phillips curve.' These are the words of economist George Akerlof from the lecture he gave when he received the Nobel Prize for Economics in 2001. The *Phillips curve* is the short-run relationship between inflation and unemployment.

Origins of the Phillips Curve

In 1958, a New Zealand economist working at the London School of Economics, A.W. Phillips, published an article in the British journal *Economica* that would make him famous. The article was entitled 'The Relationship Between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861–1957'. In it, Phillips showed a negative correlation between the rate of unemployment and the rate of inflation. That is, Phillips showed that years with low unemployment tend to have high inflation, and years with high unemployment tend to have low inflation. (It should be noted that Phillips examined inflation in nominal wages rather than inflation in prices, but for our purposes that distinction is not important. These two measures of inflation usually move together because of the relationship between wage inflation and the excess demand for labour.) Phillips concluded that two important macroeconomic variables – inflation and unemployment – were linked in a way that economists had not previously appreciated.

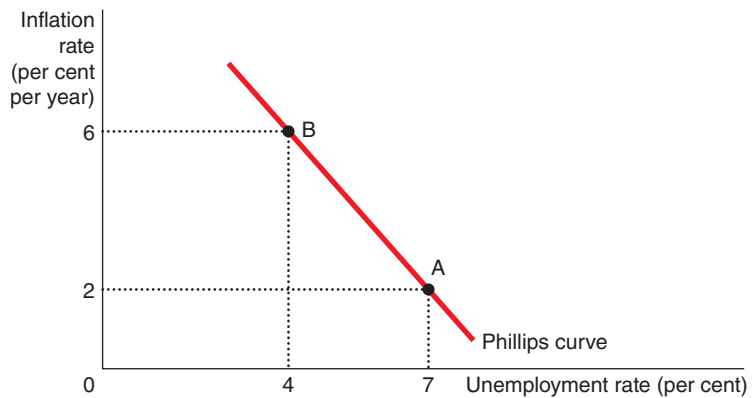
Although Phillips's discovery was based on data for the United Kingdom, researchers quickly extended his finding to other countries. Two years after Phillips published his article, economists Paul Samuelson and Robert Solow published an article in the *American Economic Review* called 'Analytics of Anti-inflation Policy' in which they showed a similar negative correlation between inflation and unemployment in data for the United States. They reasoned that this correlation arose because low unemployment was associated with high aggregate demand, which in turn puts upwards pressure on wages and prices throughout the economy. Samuelson and Solow dubbed the negative association between inflation and unemployment the **Phillips curve**. Figure 34.1 shows an example of a Phillips curve like the one found by Samuelson and Solow.

Phillips curve a curve that shows the short-run trade-off between inflation and unemployment

FIGURE 34.1

The Phillips Curve

The Phillips curve illustrates a negative association between the inflation rate and the unemployment rate. At point A, inflation is low and unemployment is high. At point B, inflation is high and unemployment is low.



As the title of their paper suggests, Samuelson and Solow were interested in the Phillips curve because they believed that it held important lessons for policymakers. In particular, they suggested that the Phillips curve offers policymakers a menu of possible economic outcomes. By altering monetary and fiscal policy to influence aggregate demand, policymakers could choose any point on this curve. Point A offers high unemployment and low inflation. Point B offers low unemployment and high inflation. Policymakers might prefer both low inflation and low unemployment, but the historical data as summarized by the Phillips curve indicate that this combination is impossible. According to Samuelson and Solow, policymakers face a trade-off between inflation and unemployment, and the Phillips curve illustrates that trade-off.

Aggregate Demand, Aggregate Supply and the Phillips Curve

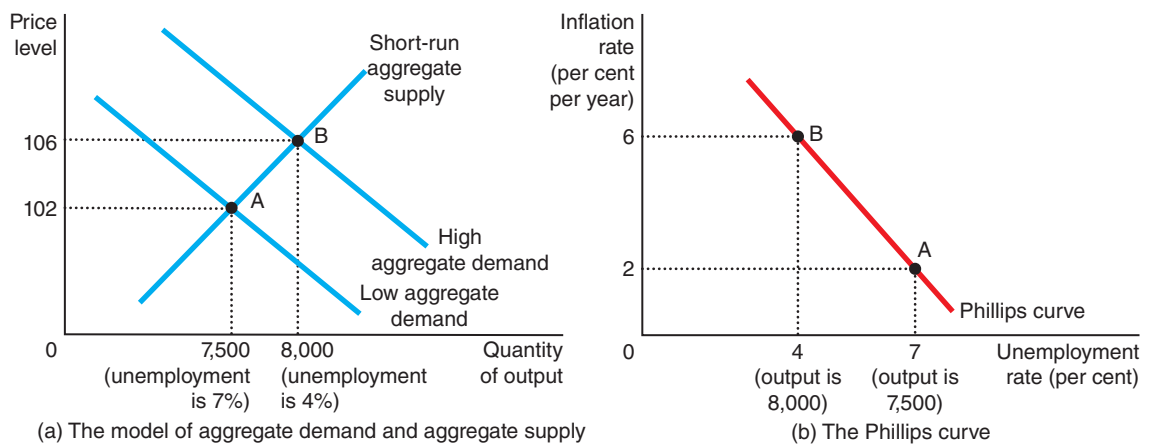
The model of aggregate demand and aggregate supply provides an easy explanation for the menu of possible outcomes described by the Phillips curve. The Phillips curve simply shows the combinations of inflation and unemployment that arise in the short run as shifts in the aggregate demand curve move the economy along the short-run aggregate supply curve. As we saw in Chapter 32, an increase in the aggregate demand for goods and services leads, in the short run, to a larger output of goods and services and a higher price level. Larger output means greater employment and, thus, a lower rate of unemployment. In addition, whatever the previous year's price level happens to be, the higher the price level in the current year, the higher the rate of inflation. Thus, shifts in aggregate demand push inflation and unemployment in opposite directions in the short run – a relationship illustrated by the Phillips curve.

To see more fully how this works, let's consider an example. To keep the numbers simple, imagine that the price level (as measured, for instance, by the consumer prices index) equals 100 in the year 2014. Figure 34.2 shows two possible outcomes that might occur in year 2015. Panel (a) shows the two outcomes using the model of aggregate demand and aggregate supply. Panel (b) illustrates the same two outcomes using the Phillips curve.

FIGURE 34.2

How the Phillips Curve is Related to the Model of Aggregate Demand and Aggregate Supply

This figure assumes a price level of 100 for the year 2014 and charts possible outcomes for the year 2015. Panel (a) shows the model of aggregate demand and aggregate supply. If aggregate demand is low, the economy is at point A; output is low (7,500), and the price level is low (102). If aggregate demand is high, the economy is at point B; output is high (8,000), and the price level is high (106). Panel (b) shows the implications for the Phillips curve. Point A, which arises when aggregate demand is low, has high unemployment (7 per cent) and low inflation (2 per cent). Point B, which arises when aggregate demand is high, has low unemployment (4 per cent) and high inflation (6 per cent).



In panel (a) of the figure, we can see the implications for output and the price level in the year 2015. If the aggregate demand for goods and services is relatively low, the economy experiences outcome A. The economy produces output of 7,500, and the price level is 102. By contrast, if aggregate demand is relatively high, the economy experiences outcome B. Output is 8,000, and the price level is 106. Thus, higher aggregate demand moves the economy to an equilibrium with higher output and a higher price level.

In panel (b) of the figure, we can see what these two possible outcomes mean for unemployment and inflation. Because firms need more workers when they produce a greater output of goods and services, unemployment is lower in outcome B than in outcome A. In this example, when output rises from 7,500 to 8,000, unemployment falls from 7 per cent to 4 per cent. Moreover, because the price level is higher at outcome B than at outcome A, the inflation rate (the percentage change in the price level from the previous year) is also higher. In particular, since the price level was 100 in year 2014, outcome A has an inflation rate of 2 per cent, and outcome B has an inflation rate of 6 per cent. Thus, we can compare the two possible outcomes for the economy either in terms of output and the price level (using the model of aggregate demand and aggregate supply) or in terms of unemployment and inflation (using the Phillips curve).

As we saw in the preceding chapter, monetary and fiscal policy can shift the aggregate demand curve. Therefore, monetary and fiscal policy can move the economy along the Phillips curve. Increases in the money supply, increases in government spending, or cuts in taxes expand aggregate demand and move the economy to a point on the Phillips curve with lower unemployment and higher inflation.

Decreases in the money supply, cuts in government spending, or increases in taxes contract aggregate demand and move the economy to a point on the Phillips curve with lower inflation and higher unemployment. In this sense, the Phillips curve offers policymakers a menu of combinations of inflation and unemployment.

SELF TEST Draw the Phillips curve. Use the model of aggregate demand and aggregate supply to show how policy can move the economy from a point on this curve with high inflation to a point with low inflation.

SHIFTS IN THE PHILLIPS CURVE: THE ROLE OF EXPECTATIONS

The Phillips curve seems to offer policymakers a menu of possible inflation–unemployment outcomes. But does this menu remain stable over time? Is the Phillips curve a relationship on which policymakers can rely? Economists took up these questions in the late 1960s, shortly after Samuelson and Solow had introduced the Phillips curve into the macroeconomic policy debate.

The Long-Run Phillips Curve

In 1968 economist Milton Friedman published a paper in the *American Economic Review*, based on an address he had recently given as president of the American Economic Association. The paper, entitled ‘The Role of Monetary Policy’, contained sections on ‘What Monetary Policy Can Do’ and ‘What Monetary Policy Cannot Do’. Friedman argued that one thing monetary policy cannot do, other than for only a short time, is pick a combination of inflation and unemployment on the Phillips curve. At about the same time, another economist, Edmund Phelps, also published a paper denying the existence of a long-run trade-off between inflation and unemployment.

Friedman and Phelps based their conclusions on classical principles of macroeconomics, which we discussed in Chapters 22 to 29. Recall that classical theory points to growth in the money supply as the primary determinant of inflation. But classical theory also states that monetary growth does not have real effects – it merely alters all prices and nominal incomes proportionately. In particular, monetary growth does not influence those factors that determine the economy’s unemployment rate, such as the market power of unions, the role of efficiency wages, or the process of job search. Friedman and Phelps concluded that there is no reason to think the rate of inflation would, *in the long run*, be related to the rate of unemployment.

Here, in his own words, is Friedman’s view about what the central bank can hope to accomplish in the long run:

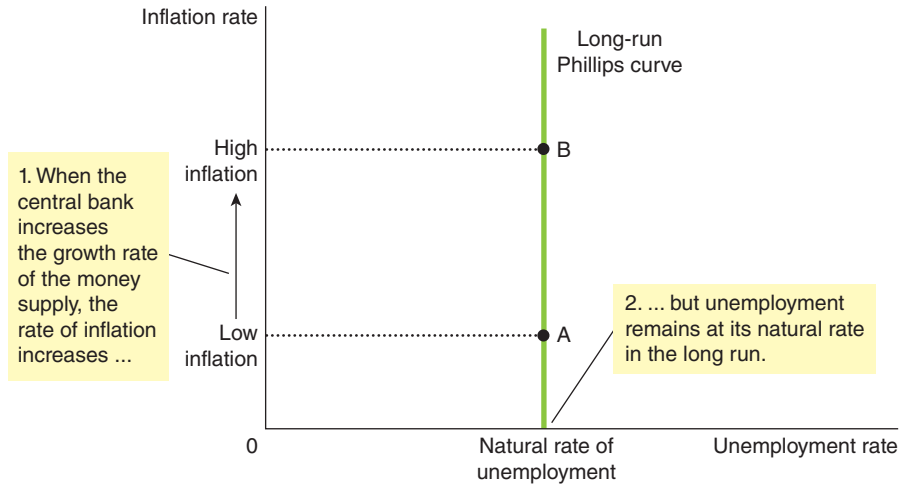
The monetary authority controls nominal quantities – directly, the quantity of its own liabilities [currency plus bank reserves]. In principle, it can use this control to peg a nominal quantity – an exchange rate, the price level, the nominal level of national income, the quantity of money by one definition or another – or to peg the change in a nominal quantity – the rate of inflation or deflation, the rate of growth or decline in nominal national income, the rate of growth of the quantity of money. It cannot use its control over nominal quantities to peg a real quantity – the real rate of interest, the rate of unemployment, the level of real national income, the real quantity of money, the rate of growth of real national income, or the rate of growth of the real quantity of money.

These views have important implications for the Phillips curve. In particular, they imply that monetary policymakers face a long-run Phillips curve that is vertical, as in Figure 34.3. If the central bank increases the money supply slowly, the inflation rate is low, and the economy finds itself at point A. If the central bank increases the money supply quickly, the inflation rate is high, and the economy finds itself at point B. In either case, the unemployment rate tends towards its normal level, called the *natural rate of unemployment*. The vertical long-run Phillips curve illustrates the conclusion that unemployment does not depend on money growth and inflation in the long run.

FIGURE 34.3

The Long-Run Phillips Curve

According to Friedman and Phelps, there is no trade-off between inflation and unemployment in the long run. Growth in the money supply determines the inflation rate. Regardless of the inflation rate, the unemployment rate gravitates towards its natural rate. As a result, the long-run Phillips curve is vertical.

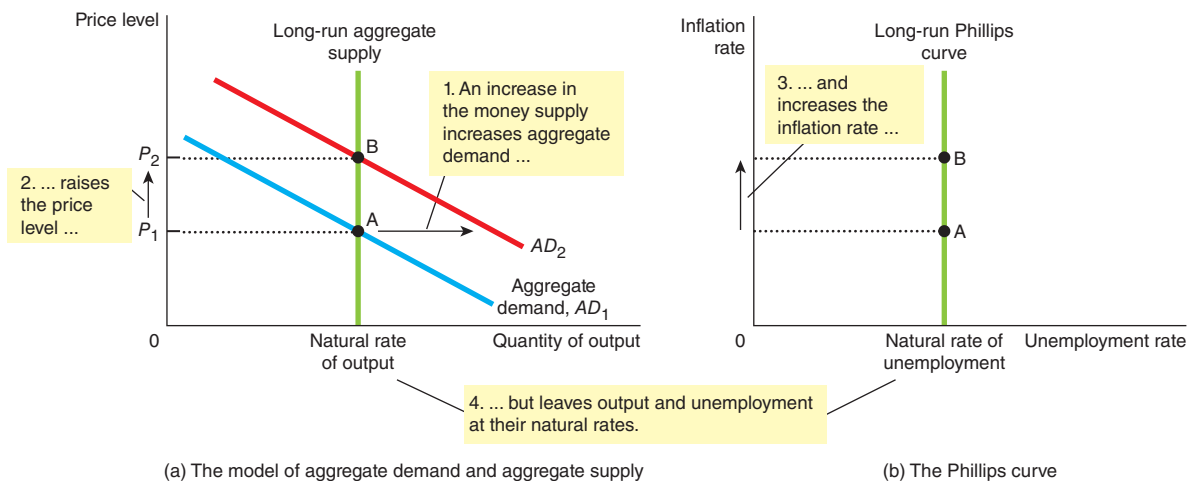


The vertical long-run Phillips curve is, in essence, one expression of the classical idea of monetary neutrality. As you may recall, we expressed this idea in Chapter 32 with a vertical long-run aggregate supply curve. Indeed, as Figure 34.4 illustrates, the vertical long-run Phillips curve and the vertical long-run aggregate supply curve are two sides of the same coin. In panel (a) of this figure, an increase in the money supply shifts the aggregate demand curve to the right from AD_1 to AD_2 . As a result of this shift, the long-run equilibrium moves from point A to point B.

FIGURE 34.4

How the Long-Run Phillips Curve Is Related to the Model of Aggregate Demand and Aggregate Supply

Panel (a) shows the model of aggregate demand and aggregate supply with a vertical aggregate supply curve. When expansionary monetary policy shifts the aggregate demand curve to the right from AD_1 to AD_2 , the equilibrium moves from point A to point B. The price level rises from P_1 to P_2 , while output remains the same. Panel (b) shows the long-run Phillips curve, which is vertical at the natural rate of unemployment. Expansionary monetary policy moves the economy from lower inflation (point A) to higher inflation (point B) without changing the rate of unemployment.



The price level rises from P_1 to P_2 , but because the aggregate supply curve is vertical, output remains the same. In panel (b), more rapid growth in the money supply raises the inflation rate by moving the economy from point A to point B. But, because the Phillips curve is vertical, the rate of unemployment is

the same at these two points. Thus, the vertical long-run aggregate supply curve and the vertical long-run Phillips curve both imply that monetary policy influences nominal variables (the price level and the inflation rate) but not real variables (output and unemployment). Regardless of the monetary policy pursued by the central bank, output and unemployment are, in the long run, at their natural rates.

The Natural Rate of Unemployment What is so ‘natural’ about the natural rate of unemployment? Friedman and Phelps used this adjective to describe the unemployment rate towards which the economy tends to gravitate in the long run. Yet the natural rate of unemployment is not necessarily the socially desirable rate of unemployment. Nor is the natural rate of unemployment constant over time. For example, suppose that a newly formed union uses its market power to raise the real wages of some workers above the equilibrium level. The result is an excess supply of workers and, therefore, a higher natural rate of unemployment. This unemployment is ‘natural’ not because it is good but because it is beyond the influence of monetary policy. More rapid money growth would not reduce the market power of the union or the level of unemployment; it would lead only to more inflation.

Although monetary policy cannot influence the natural rate of unemployment, other types of policy can. To reduce the natural rate of unemployment, policymakers should look to policies that improve the functioning of the labour market. Earlier in the book we discussed how various labour market policies, such as minimum wage laws, collective bargaining laws, unemployment insurance and job-training schemes, affect the natural rate of unemployment. A policy change that reduced the natural rate of unemployment would shift the long-run Phillips curve to the left. In addition, because lower unemployment means more workers are producing goods and services, the quantity of goods and services supplied would be larger at any given price level, and the long-run aggregate supply curve would shift to the right. The economy could then enjoy lower unemployment and higher output for any given rate of money growth and inflation. We will look at some of these policies designed to affect the supply side of the economy in the next chapter.

CASE STUDY A Fear of Inflation

A key debate which continues following the financial crisis is whether weak growth being experienced in many countries needs to be countered by expansionary fiscal and monetary policy, and if so what danger there is to inflation in the coming years. There is a body of opinion that the sort of expansion necessary would lead to severe inflationary consequences and that expectations on managing inflation at stable levels need to be maintained. Such an argument would embrace the basic premise of the Phillips curve.

Some economists argue that monetary and fiscal expansion at this time would not lead to higher inflation because the economies of many countries are so depressed that there is no evidence to suggest that inflationary pressures exist outside of the increase in fuel prices that have been a major cause of higher than target inflation in countries like the UK. This would suggest that the size of the output or deflationary gap is sufficiently large to accommodate expansion without leading to inflationary pressures. This, in turn, suggests that shape of the Phillips curve would be such that the reduction in unemployment would be larger in comparison to the rise in inflation (the Phillips curve as we have drawn it in Figure 34.1 would be shallow).



Rising fuel prices have meant inflation has stayed stubbornly high during the economic slowdown in some countries but would monetary expansion lead to the release of inflationary pressure across the economy as a whole?

Reconciling Theory and Evidence

At first, the conclusion of Friedman and Phelps of a long-run trade-off between inflation and unemployment might not seem persuasive. Their argument was based on an appeal to *theory*. In contrast, the negative correlation between inflation and unemployment documented by Phillips, Samuelson and Solow was based on *data*. Why should anyone believe that policymakers faced a vertical Phillips curve when the world seemed to offer a downwards sloping one? Shouldn't the findings of Phillips, Samuelson and Solow lead us to reject the classical conclusion of monetary neutrality?

Friedman and Phelps were well aware of these questions, and they offered a way to reconcile classical macroeconomic theory with the finding of a downwards sloping Phillips curve in data from the United Kingdom and the United States. They claimed that a negative relationship between inflation and unemployment holds in the short run but that it cannot be used by policymakers in the long run. In other words, policymakers can pursue expansionary monetary policy to achieve lower unemployment for a while, but eventually unemployment returns to its natural rate, and more expansionary monetary policy leads only to higher inflation.

Friedman and Phelps reasoned as we did in Chapter 32 when we explained the difference between the short-run and long-run aggregate supply curves. (In fact, the discussion in that chapter drew heavily on the legacy of Friedman and Phelps.) As you may recall, the short-run aggregate supply curve is upwards sloping, indicating that an increase in the price level raises the quantity of goods and services that firms supply. In contrast, the long-run aggregate supply curve is vertical, indicating that the price level does not influence quantity supplied in the long run. Chapter 32 presented three theories to explain the upwards slope of the short-run aggregate supply curve: sticky wages, sticky prices and misperceptions about relative prices. Because wages, prices and perceptions adjust to changing economic conditions over time, the positive relationship between the price level and quantity supplied applies in the short run but not in the long run. Friedman and Phelps applied this same logic to the Phillips curve. Just as the aggregate supply curve slopes upwards only in the short run, the trade-off between inflation and unemployment holds only in the short run. And just as the long-run aggregate supply curve is vertical, the long-run Phillips curve is also vertical.

To help explain the short-run and long-run relationship between inflation and unemployment, Friedman and Phelps introduced a new variable into the analysis: *expected inflation*. Expected inflation measures how much people expect the overall price level to change. As we discussed in Chapter 32, the expected price level affects the wages and prices that people set and the perceptions of relative prices that they form. As a result, expected inflation is one factor that determines the position of the short-run aggregate supply curve. In the short run, the central bank can take expected inflation (and thus the short-run aggregate supply curve) as already determined. When the money supply changes, the aggregate demand curve shifts and the economy moves along a given short-run aggregate supply curve. In the short run, therefore, monetary changes lead to unexpected fluctuations in output, prices, unemployment and inflation. In this way, Friedman and Phelps explained the Phillips curve that Phillips, Samuelson and Solow had documented.

Yet the central bank's ability to create unexpected inflation by increasing the money supply exists only in the short run. In the long run, people come to expect whatever inflation rate the central bank chooses to produce. Because wages, prices and perceptions will eventually adjust to the inflation rate, the long-run aggregate supply curve is vertical. In this case, changes in aggregate demand, such as those due to changes in the money supply, do not affect the economy's output of goods and services. Thus, Friedman and Phelps concluded that unemployment returns to its natural rate in the long run.

The Short-Run Phillips Curve

The analysis of Friedman and Phelps can be summarized in the following equation (which is, in essence, another expression of the aggregate supply equation we saw in Chapter 32):

$$\text{Unemployment rate} = \text{Natural rate of unemployment} - a(\text{Actual Inflation} - \text{Expected inflation})$$

This equation relates the unemployment rate to the natural rate of unemployment, actual inflation and expected inflation. In the short run, expected inflation is given. As a result, higher actual inflation is

associated with lower unemployment. (How much unemployment responds to unexpected inflation is determined by the size of a , a number that in turn depends on the slope of the short-run aggregate supply curve.) In the long run, however, people come to expect whatever inflation the central bank produces. Thus, actual inflation equals expected inflation, and unemployment is at its natural rate.

This equation implies there is no stable short-run Phillips curve. Each short-run Phillips curve reflects a particular expected rate of inflation. (To be precise, if you graph the equation, you'll find that the short-run Phillips curve intersects the long-run Phillips curve at the expected rate of inflation.) Whenever expected inflation changes, the short-run Phillips curve shifts.

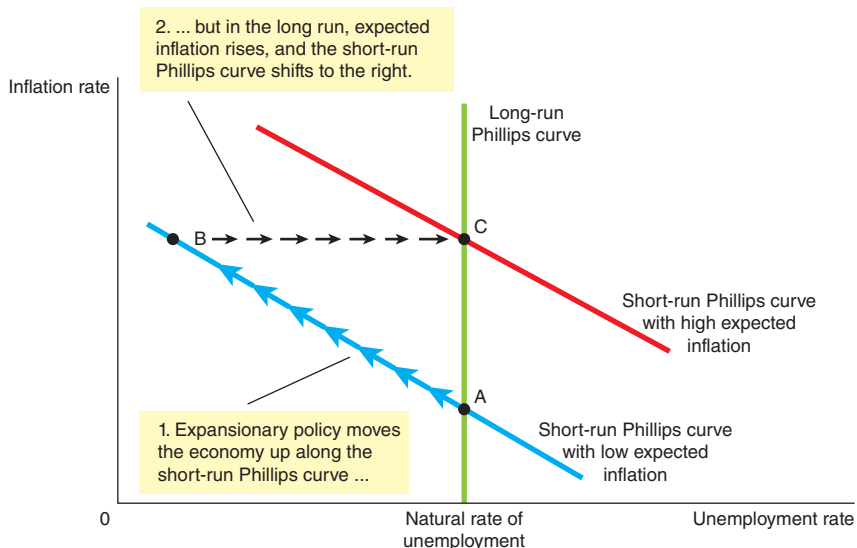
According to Friedman and Phelps, it is dangerous to view the Phillips curve as a menu of options available to policymakers. To see why, imagine an economy at its natural rate of unemployment with low inflation and low expected inflation, shown in Figure 34.5 as point A. Now suppose that policymakers try to take advantage of the trade-off between inflation and unemployment by using monetary or fiscal policy to expand aggregate demand. In the short run when expected inflation is given, the economy goes from point A to point B. Unemployment falls below its natural rate, and inflation rises above expected inflation. Over time, people get used to this higher inflation rate, and they raise their expectations of inflation. When expected inflation rises, firms and workers start taking higher inflation into account when setting wages and prices. The short-run Phillips curve then shifts to the right, as shown in the figure. The economy ends up at point C, with higher inflation than at point A but with the same level of unemployment.

Thus, Friedman and Phelps concluded that policymakers do face a trade-off between inflation and unemployment, but only a temporary one. If policymakers use this trade-off, they lose it.

FIGURE 34.5

How Expected Inflation Shifts the Short-Run Phillips Curve

The higher the expected rate of inflation, the higher the short-run trade-off between inflation and unemployment. At point A, expected inflation and actual inflation are both low, and unemployment is at its natural rate. If the central bank pursues an expansionary monetary policy, the economy moves from point A to point B in the short run. At point B, expected inflation is still low, but actual inflation is high. Unemployment is below its natural rate. In the long run, expected inflation rises, and the economy moves to point C. At point C, expected inflation and actual inflation are both high, and unemployment is back to its natural rate.



The Unemployment–Inflation Trade-Off

Friedman and Phelps had made a bold prediction in 1968: if policymakers try to take advantage of the Phillips curve by choosing higher inflation in order to reduce unemployment, they will succeed at reducing

unemployment only temporarily. This view – that unemployment eventually returns to its natural rate, regardless of the rate of inflation – is called the **natural-rate hypothesis**.

natural-rate hypothesis the claim that unemployment eventually returns to its normal, or natural, rate, regardless of the rate of inflation

To some economists at the time, it seemed ridiculous to claim that the Phillips curve would break down once policymakers tried to use it. But, in fact, that is exactly what happened in both the UK and the United States. Beginning in the late 1960s, the UK government, for example, followed policies that expanded the aggregate demand for goods and services. On top of this, the UK and many other developed economies in the late 1960s and early 1970s experienced an increase in aggregate demand due to American involvement in the Vietnam War, which increased US government spending (on the military), boosted US aggregate demand and so boosted net exports from other countries to the USA. In addition, in 1971, as a result of the relaxation of certain controls on bank lending, the UK experienced a major expansion in the money supply. In the following year, the government announced an extraordinarily expansionary fiscal policy, in terms of extra spending and tax reduction, and the economy began seriously to overheat and inflation started to rise (refer to Figure 30.2 to see the extent to which the economy of the UK grew at this time). But, as Friedman and Phelps had predicted, unemployment did not stay low.

From the end of World War II through to the late 1970s, many governments in the Western world adopted Keynesian policies as the basis for managing the economy. There was an emphasis on attempting to keep unemployment low. The experience of mass unemployment and the rise in extremism in many parts of Europe and the misery caused by the Great Depression in the United States in the 1930s had persuaded policymakers that a focus on unemployment was vital to stable economic and political life. In the 20 years after the war, fiscal policy was dominant; in times of rising unemployment fiscal policy was loosened and tightened when inflation began to speed up. The government could pull its fiscal levers of tax and public spending to ‘fine tune’ the economy. The unusual circumstances of post-war reconstruction across Europe and the US meant that inflation did not seem to present a major problem.

Some economists argued that the focus on fiscal policy meant that not only did households form expectations about inflation but also about government policy. In the UK, many large industries were in public ownership including the railways, coal, electricity, airports and airlines, telecommunications and steel. In addition, trade unions enjoyed considerable power in many industries. Workers came to expect, it was argued, increases in real wages each year and that this would be accommodated by monetary expansion. In addition, the government would respond by loosening fiscal policy if the economy slowed down. This, it was argued, led to a fall in productivity and increased inflationary pressures as the government and industry accommodated wage increases. In the aggregate, UK industry became sluggish, lacked flexibility and lost competitiveness.

These conditions led to rising inflation and an inability of the government to reduce unemployment below the natural rate. After the oil crisis and miners’ strike of the early 1970s and the fiscal boom produced by the government, stagflation took hold and the UK seemed to lurch from ‘boom to bust’. Eventually the government was forced to borrow money from the International Monetary Fund (IMF) in 1976. In a famous speech to the Labour Party Conference in the same year, the then Prime Minister, James Callaghan, echoed Friedman and Phelps’ theory to the nation when he said:

We used to think that you could spend your way out of a recession and increase employment by cutting taxes and boosting government spending. I tell you in all candour that that option no longer exists, and in so far as it ever did exist, it only worked on each occasion since the war by injecting a bigger dose of inflation into the economy. And each time that happened, the average level of unemployment has risen. Higher inflation followed by higher unemployment. That is the history of the last 20 years.

This effectively saw a move away from Keynesian policies and a shift to monetary policy in the UK, while the US began to take greater steps to focus on the money supply. The focus of policy on the use of monetary and supply-side policies was seen as the way of keeping inflation under control and lowering the

natural rate of unemployment. Throughout the 1980s, the UK and US governments trumpeted the benefits of supply-side policies. These looked at ways of expanding the productive capacity of the economy to shift the aggregate supply curve to the right (we will look at these policies in the next chapter). Such policies stressed the importance of enterprise, reducing business regulation, improving incentives through cutting taxes and benefits, reducing trade union power and investing in education and training to improve the workings and flexibility of the labour market in the long run.

Supply-side policies take some time to have an effect but structural economic reforms by the UK government throughout the 1990s led to a leftward shift in the Phillips curve. What caused this favourable shift in the short-run Phillips curve? Part of the answer lies in a low level of expected inflation. The policy of inflation targeting since 1992, combined with the independence of the Bank of England in 1997 created a credible policy framework in which workers and firms knew that interest rates would be raised if the economy began to overheat, and they tended to moderate their wage claims and price setting accordingly. Since the Bank of England now sets interest rates independently, people also know that there is no way in which politicians can use expansionary monetary policy for political reasons, such as to gain popularity before an election. This led to a period of relatively stable economic conditions in the UK with economic growth, low inflation and relatively low levels of unemployment right up until around 2008 when the financial crisis began to unravel.

In contrast, many European countries suffered relatively high levels of unemployment despite the European Central Bank having independence in the setting of monetary policy. Why was this? One argument that has been put forward to explain high European unemployment in the 21st century, and which seems to have some credibility, centres on the level of labour market regulation. As we pointed out above, the 1980s in the UK were characterized both by a weakening of the power of the trades unions and by a reduction in business regulation in general and in labour market regulation in particular. Whilst, for example, this had the effect of reducing job security for many people by making it easier for employers to terminate contracts, it also had the effect of making labour markets much more flexible. Thus, somewhat paradoxically perhaps, if it is harder to fire someone, firms will think hard before taking on new labour and unemployment may actually rise. Similarly, the minimum wage is typically set at a much higher level (relative to the average wage) in European countries like France and Germany, which again is a reason why the natural rate of unemployment might be higher in those countries.

Data on the history of UK inflation and unemployment from the early 1970s shows very little evidence of the simple negative relationship between these two variables that Phillips had originally observed. In particular, as inflation remained high in the early 1970s, people's expectations of inflation caught up with reality, and the unemployment rate rose along with inflation. By the mid-1970s, policymakers had learned that Friedman and Phelps were right: There is no trade-off between inflation and unemployment in the long run.

SELF TEST Draw the short-run Phillips curve and the long-run Phillips curve. Explain why they are different.

THE LONG-RUN VERTICAL PHILLIPS CURVE AS AN ARGUMENT FOR CENTRAL BANK INDEPENDENCE

Who would you rather was in charge of monetary policy – an elected government or an unelected central banker? Would it not be better to have the government in charge of the money supply and interest rates? After all, if they screw up, we can vote them out of office at the next election, while, on the other hand, the electorate has no such sanction over central bankers.

In fact, most economists – at least the majority of those who believe in a long-run vertical Phillips curve – would rather hand control of monetary policy to the central banker, particularly if the central banker in question had a reputation for being tough on inflation, and especially if he or she had a clear and publicly known target for the level of inflation to be achieved.

To see why, take another look back at Figure 34.5. Suppose we're at point A, with the natural rate of unemployment and a low level of inflation, and that the government is in charge of monetary policy in the sense that they can just tell the central bank to change interest rates or the money supply and it is done. Now imagine that we're just coming up to a general election in a few months' time: what do you suppose the government will do? Since unemployment is never a vote winner, it is perhaps likely that they will be tempted to reduce interest rates and expand the money supply so that the economy moves from point A to point B and unemployment falls. (In reality, because of the uncertain lags in monetary policy that we discussed earlier, the government won't be able to guarantee that the economy will get to precisely point B just before the election, but it will certainly move along the short-run Phillips curve from point A towards point B, with actual inflation above expected inflation.) Of course, once people's expectations of inflation catch up with actual inflation, the short-run Phillips curve will shift up and the economy will move to point C. At point C there is the same rate of unemployment as in the first place (the natural rate) and a higher level of inflation, so in the long run the economy is in a worse state than if monetary policy had remained neutral and the economy had stayed at A. If the election takes place before inflationary expectations catch up, however, the government has a higher chance of being re-elected because, so long as the economy remains on the initial short-run Phillips curve at some point between A and B, inflationary expectations are low while unemployment is falling. Not only is this likely to lead to extra votes from people who were previously unemployed and now grateful to be in a job, economic activity in general will rise in the economy and people in general will experience an economic 'feel-good factor' and be grateful to the government for running the economy so well. Once inflationary expectations have finally caught up and we move to point C on the long-run Phillips curve with the same old level of unemployment and higher inflation, the feel-good factor will evaporate, but it's too late – the election is over and the same government is back in power.

Of course, the story we have just told implies a pretty cynical view of politics, and most governments would strenuously deny that they would carry out such tricks just to get re-elected. More importantly, however, the story also implies that people (workers and managers of firms) are pretty stupid. Why? Because everyone will know that the government has a strong incentive to pursue an expansionary monetary policy just before an election. In fact, firms and workers may begin raising prices and wages before the election in anticipation of the expansionary monetary policy, so we could end up jumping straight from point A to point C without any intervening fall in unemployment! This is an interesting result, and one worth stressing: if people believe that the government is about to pursue an expansionary monetary policy, then inflationary expectations will increase and inflation will rise but unemployment will not fall. In essence, the economy jumps immediately to the new long-run equilibrium.

But suppose that whoever is in charge of monetary policy issues a statement saying that she has no intention of pursuing an expansionary monetary policy. Should firms and workers believe her? Probably not. In fact, firms and workers might reason that if they really believed that no monetary expansion would take place, then that is when the politician would be most tempted to bring about an expansion of the money supply, since then they really would be taken by surprise by the monetary expansion and it would take a while for their inflationary expectations to catch up with actual inflation. In the meantime, the economy would move along the short-run Phillips curve to point B, unemployment would fall and the government would become more popular. So the more credible the politician's promise seems to be, the less you should believe her! Hence, it seems that if monetary policy is left to the politicians, we shall always end up with higher inflation but with no reduction in unemployment, even in the short run.

What if, instead, the government gave the governor of the central bank full control over monetary policy and told him that he would be fired unless he achieved a certain low level of inflation over a reasonable period of time, regardless of what happened to unemployment? Say, in fact, that this target level of inflation was the one at point A in Figure 34.5. Because the central banker has no incentive to reduce unemployment, the problem goes away: firms and workers believe that the economy will stay at point A and so it does, because inflationary expectations are exactly the same as actual inflation. Hence, by passing control over monetary policy to the central bank, the economy ends up with unemployment no different from when the government was in control of monetary policy, but with a lower level of inflation.

So what's in it for the government? Well, for a start we should not perhaps think that all politicians are necessarily as cynical as we have portrayed them. Perhaps they have the longer-run well-being of the economy at heart rather than just wanting to be re-elected at any cost. Or perhaps they are cynical, but just more sophisticated. A cynical government would clearly prefer point B (or somewhere near it) to

point A at election time, since unemployment is lower than the natural rate and people still expect low inflation – a sure vote-winning combination. However, the government knows that firms and workers will factor the government’s temptation into their price and wage setting, so that in fact the economy will jump to point C just before the election.

In fact, this situation is an example of a *Nash equilibrium*. Remember that a Nash equilibrium is a situation in which economic actors interacting with one another each choose their best strategy given the strategies that all the other actors have chosen. Here, firms and workers know that if they don’t raise inflationary expectations before an election, the government will most likely pursue an expansionary monetary policy and they will lose out because their prices and wages will be lower in real terms, and so they do raise inflationary expectations and the short-run Phillips curve shifts up and we move to point C in Figure 34.5. But doesn’t the government still have the same temptation to inflate at point C? Well, possibly, but there will certainly be some point on the vertical, long-run Phillips curve where inflation is already so high that the government will not want to risk pushing it higher, even if it means a short-run reduction in unemployment. If firms and workers can guess roughly what this level of inflation is, they will set their wages and prices so that the economy will jump straight to that point – let’s imagine that this is indeed point C. If, at point C, the government has no temptation to inflate the economy in order to gain popularity, and if firms and workers know this so that they have no incentive to change their price and wage setting behaviour, then we will have reached the Nash equilibrium.

Another way of thinking about this is to say that the Nash equilibrium (which we have assumed is at point C in Figure 34.5) represents the *time-consistent policy*. A time-consistent policy is simply one which a government does not have a temptation to renege on at some point in time and will usually represent a Nash equilibrium.

But take a look at point C: it may represent time-consistency in economic policy, but it is actually worse than point A for the government’s electoral chances, since inflation is higher and unemployment is still at the natural rate. Hence, the best thing to do in order to maximize the chances of re-election is for the government to make the central bank independent in the sense of handing over control of monetary policy – providing of course that the central bank sees its role as the guardian of price stability; in other words, providing that the central bank is ‘conservative’ with respect to price stability. That way, the economy can just stay at A.

It is a testimony to the power of macroeconomic theory that this argument has persuaded many governments around the world to grant independence to their central bank in the conduct of monetary policy. The European Central Bank, for example, has been independent since its inception in 1998, and the Bank of England was granted independence in 1997. The ECB both designs its monetary policy (e.g. decides what level of inflation in the Euro Area to aim for) and implements it. The Bank of England, on the other hand, has independence in the implementation of monetary policy but does not decide on the design of monetary policy – in particular, its inflation target is set by the UK Chancellor of the Exchequer. The US Federal Reserve is also independent in both the design and implementation of monetary policy. Although, unlike the ECB and the Bank of England, the Fed does not formally pursue an inflation target, it has a strong reputation for ‘conservatism’ with respect to inflation, so the overall effect is similar. A long-run vertical Phillips curve is a compelling case for taking control over monetary policy out of the hands of politicians and handing it over to a ‘conservative central banker’.

SHIFTS IN THE PHILLIPS CURVE: THE ROLE OF SUPPLY SHOCKS

Friedman and Phelps had suggested in 1968 that changes in expected inflation shift the short-run Phillips curve, and the experience of the early 1970s convinced most economists that Friedman and Phelps were right. Within a few years, however, the economics profession would turn its attention to a different source of shifts in the short-run Phillips curve: shocks to aggregate supply.

This time, the shift in focus came not from two economics professors but from a group of Arab sheikhs. Conflict between Israel and its Arab neighbours triggered a series of oil price shocks as Arab oil producers used their market power to exert political pressure on Western governments who supported Israel. In 1974, the Organization of Petroleum Exporting Countries (OPEC) also began to exert its power as a cartel

in order to increase its members' profits. The countries of OPEC, such as Saudi Arabia, Kuwait and Iraq, restricted the amount of crude oil they pumped and sold on world markets. This reduction in supply caused the price of oil to almost double over a few years in the 1970s.

A large increase in the world price of oil is an example of a supply shock. A **supply shock** is an event that directly affects firms' costs of production and thus the prices they charge; it shifts the economy's aggregate supply curve and, as a result, the Phillips curve. Oil is a constituent part of so many production processes that increases in its price have far-reaching effects. For example, when an oil price increase raises the cost of producing petrol, heating oil, tyres, plastic products, distribution and many other products, it reduces the quantity of goods and services supplied at any given price level. As panel (a) of Figure 34.6 shows, this reduction in supply is represented by the leftward shift in the aggregate supply curve from AS_1 to AS_2 . The price level rises from P_1 to P_2 , and output falls from Y_1 to Y_2 and the economy experiences stagflation.

supply shock an event that directly alters firms' costs and prices, shifting the economy's aggregate supply curve and thus the Phillips curve

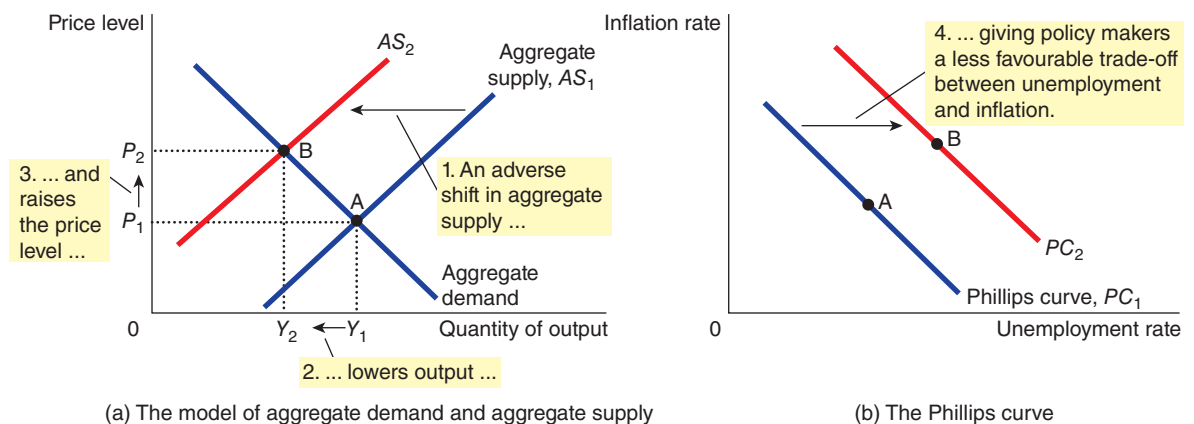
This shift in aggregate supply is associated with a similar shift in the short-run Phillips curve, shown in panel (b). Because firms need fewer workers to produce the smaller output, employment falls and unemployment rises. Because the price level is higher, the inflation rate – the percentage change in the price level from the previous year – is also higher. Thus, the shift in aggregate supply leads to higher unemployment and higher inflation. The short-run trade-off between inflation and unemployment shifts to the right from PC_1 to PC_2 .

Confronted with an adverse shift in aggregate supply, policymakers face a difficult choice between fighting inflation and fighting unemployment. If they contract aggregate demand to fight inflation, they will raise unemployment further. If they expand aggregate demand to fight unemployment, they will raise inflation further. In other words, policymakers face a less favourable trade-off between inflation and unemployment than they did before the shift in aggregate supply: they have to live with a higher rate of inflation for a given rate of unemployment, a higher rate of unemployment for a given rate of inflation, or some combination of higher unemployment and higher inflation.

FIGURE 34.6

An Adverse Shock to Aggregate Supply

Panel (a) shows the model of aggregate demand and aggregate supply. When the aggregate supply curve shifts to the left from AS_1 to AS_2 , the equilibrium moves from point A to point B. Output falls from Y_1 to Y_2 , and the price level rises from P_1 to P_2 . Panel (b) shows the short-run trade-off between inflation and unemployment. The adverse shift in aggregate supply moves the economy from a point with lower unemployment and lower inflation (point A) to a point with higher unemployment and higher inflation (point B). The short-run Phillips curve shifts to the right from PC_1 to PC_2 . Policymakers now face a worse trade-off between inflation and unemployment.



An important question is whether this adverse shift in the Phillips curve is temporary or permanent. The answer depends on how people adjust their expectations of inflation. If people view the rise in inflation due to the supply shock as a temporary aberration, expected inflation does not change, and the Phillips curve will soon revert to its former position. But if people believe the shock will lead to a new era of higher inflation, then expected inflation rises, and the Phillips curve remains at its new, less desirable position.

SELF TEST Give an example of a favourable shock to aggregate supply. Use the model of aggregate demand and aggregate supply to explain the effects of such a shock. How does it affect the Phillips curve?

THE COST OF REDUCING INFLATION

In 1979, the UK electorate voted in Margaret Thatcher as Prime Minister. The Conservative government she headed pursued policies that were focused on the tenets of monetarism and on the supply side – policies which were effectively rooted in the idea of a vertical long-run supply curve or vertical long-run Phillips curve. With inflation running at over 20 per cent and unemployment rising, Mrs Thatcher's government focused their attention on a policy of *disinflation* – a reduction in the rate of inflation. Economic theory suggested how this could be done through controlling the money supply. But what would be the short-run cost of disinflation? The answer to this question was much less certain.

The Sacrifice Ratio

To reduce the inflation rate, the central bank has to pursue contractionary monetary policy. Figure 34.7 shows some of the effects of such a decision. When the central bank slows the rate at which the money supply is growing, it contracts aggregate demand. The fall in aggregate demand, in turn, reduces the quantity of goods and services that firms produce, and this fall in production leads to a fall in employment. The economy begins at point A in the figure and moves along the short-run Phillips curve to point B, which has lower inflation and higher unemployment. Over time, as people come to understand that prices are rising more slowly, expected inflation falls, and the short-run Phillips curve shifts downward. The economy moves from point B to point C. Inflation is lower, and unemployment is back at its natural rate.

Thus, if a nation wants to reduce inflation, it must endure a period of high unemployment and low output. In Figure 34.7, this cost is represented by the movement of the economy through point B as it travels from point A to point C. The size of this cost depends on the slope of the Phillips curve and how quickly expectations of inflation adjust to the new monetary policy.

Many studies have examined the data on inflation and unemployment in order to estimate the cost of reducing inflation. The findings of these studies are often summarized in a statistic called the **sacrifice ratio**. The sacrifice ratio is the number of percentage points of annual output lost in the process of reducing inflation by 1 percentage point. A typical estimate of the sacrifice ratio is around 3 to 5. That is, for each percentage point that inflation is reduced, 3 to 5 per cent of annual output must be sacrificed in the transition.

sacrifice ratio the number of percentage points of annual output lost in the process of reducing inflation by 1 percentage point

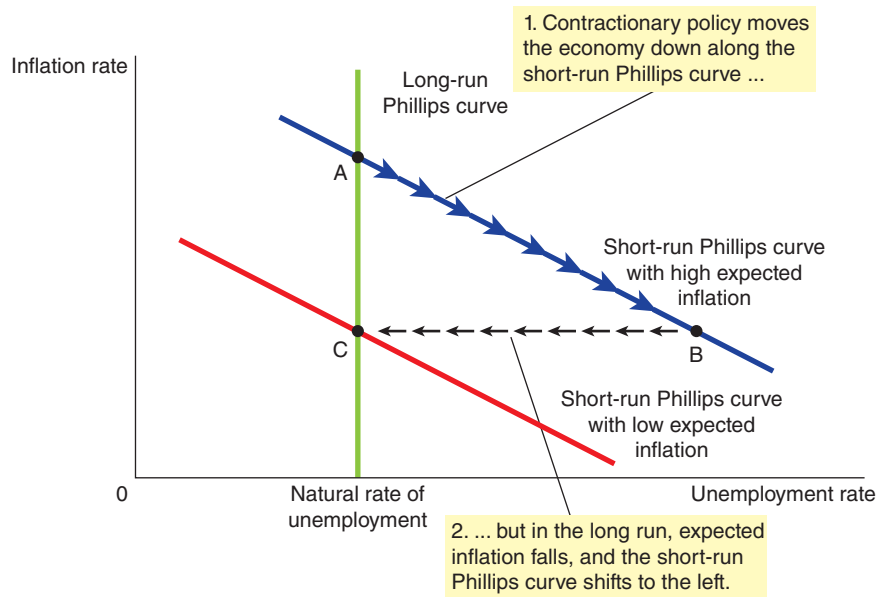
According to studies of the Phillips curve and the cost of disinflation, this sacrifice could be paid in various ways. Assume that inflation was running at 10 per cent and a government wanted to reduce it to 5 per cent. If each percentage point reduction in inflation would cost 3 per cent of annual output then the

cost would be 15 per cent of annual output lost in a year. Such an outcome would be extremely harsh even for the most hard-line inflation hawks. Another option would be to spread out the cost over several years. If the reduction in inflation took place over five years, for instance, then output would have to average only 3 per cent below trend during that period to add up to a sacrifice of 15 per cent. An even more gradual approach would be to reduce inflation slowly over a decade. Whatever path was chosen, however, reducing inflation would not be easy.

FIGURE 34.7

Disinflationary Monetary Policy in the Short Run and Long Run

When the central bank pursues contractionary monetary policy to reduce inflation, the economy moves along a short-run Phillips curve from point A to point B. Over time, expected inflation falls, and the short-run Phillips curve shifts downwards. When the economy reaches point C, unemployment is back at its natural rate.



Rational Expectations and the Possibility of Costless Disinflation

As policymakers in the early 1980s were pondering how costly reducing inflation might be, a group of economics professors were leading an intellectual revolution that would challenge the conventional wisdom on the sacrifice ratio. This group included such prominent economists as Robert Lucas, Thomas Sargent and Robert Barro. Their revolution was based on a new approach to economic theory and policy called **rational expectations**.

rational expectations the theory according to which people optimally use all the information they have, including information about government policies, when forecasting the future

The rational expectations model came to supplant the **adaptive expectations** model which was the idea that individuals and organizations base their expectations of inflation in the future on past actual inflation rates. Critics of the model argued that individuals take into account more information when forming their judgement than simply past inflation rates. The rational expectations model took into account this

available information. According to the theory of rational expectations, therefore, people optimally use all the information they have, including information about government policies, when forecasting the future.

adaptive expectations a model which states that individuals and organizations base their expectations of inflation in the future on past actual inflation rates

This new approach has had profound implications for many areas of macroeconomics, but none is more important than its application to the trade-off between inflation and unemployment. As Friedman and Phelps had first emphasized, expected inflation is an important variable that explains why there is a trade-off between inflation and unemployment in the short run but not in the long run. How quickly the short-run trade-off disappears depends on how quickly expectations adjust. Proponents of rational expectations built on the Friedman–Phelps analysis argue that when economic policies change, people adjust their expectations of inflation accordingly. Studies of inflation and unemployment that tried to estimate the sacrifice ratio had failed to take account of the direct effect of the policy regime on expectations. As a result, estimates of the sacrifice ratio were, according to the rational expectations theorists, unreliable guides for policy.

In a 1982 paper entitled ‘The End of Four Big Inflations’ (one of which was the UK inflation of the late 1970s and early 1980s), Thomas Sargent described this new view as follows:

An alternative ‘rational expectations’ view denies that there is any inherent momentum to the present process of inflation. This view maintains that firms and workers have now come to expect high rates of inflation in the future and that they strike inflationary bargains in light of these expectations. However, it is held that people expect high rates of inflation in the future precisely because the government’s current and prospective monetary and fiscal policies warrant those expectations. ... An implication of this view is that inflation can be stopped much more quickly than advocates of the ‘momentum’ view have indicated and that their estimates of the length of time and the costs of stopping inflation in terms of forgone output are erroneous. This is not to say that it would be easy to eradicate inflation. On the contrary, it would require more than a few temporary restrictive fiscal and monetary actions. It would require a change in the policy regime. ... How costly such a move would be in terms of forgone output and how long it would be in taking effect would depend partly on how resolute and evident the government’s commitment was.

According to Sargent, the sacrifice ratio could be much smaller than suggested by previous estimates. Indeed, in the most extreme case, it could be zero. If the government made a credible commitment to a policy of low inflation, people would be rational enough to lower their expectations of inflation immediately. The short-run Phillips curve would shift downward, and the economy would reach low inflation quickly without the cost of temporarily high unemployment and low output. The credibility of government policy is thus of prime importance.

The Thatcher Disinflation

In the UK, Prime Minister Margaret Thatcher did succeed at reducing inflation. Inflation came down from almost 20 per cent in 1980 to about 5 per cent in 1983 and 1984. The Thatcher disinflation did, however, come at the cost of high unemployment. In 1982 and 1983, the unemployment rate was about 11 per cent – about double its level when the Thatcher government came to power. At the same time, the production of goods and services as measured by real GDP was well below its trend level.

Does this experience refute the possibility of costless disinflation as suggested by the rational expectations theorists? Some economists have argued that the answer to this question is a resounding yes. To make the transition from high inflation (point A in Figure 34.7) to low inflation (point C), the economy had to experience a painful period of high unemployment (point B).

Yet perhaps there is good reason not to reject the conclusions of the rational expectations theorists so quickly. Even though the Thatcher disinflation did impose a cost of temporarily high unemployment, the

cost was not as large as many economists had predicted. Most estimates of the sacrifice ratio based on the Thatcher disinflation are smaller than estimates that had been obtained from previous data. It seems that Prime Minister Thatcher's tough stand on inflation did have some direct effect on expectations, as the rational expectations theorists claimed.

SELF TEST What is the sacrifice ratio? How might the credibility of the government's commitment to reduce inflation affect the sacrifice ratio?

FYI



Wage Curve Theory

We have said that Phillips's research looked at the relationship between the rate of growth of wages and unemployment. The relationship of the wage rate to unemployment takes us a small step towards linking the level of inflation with unemployment. The rate of growth in wages is a factor that influences the overall level of prices in the economy. Wages represent a cost to employers; other things being equal, if they rise then employers might well seek to raise prices to cover the increased cost. If this happens throughout the economy on an aggregate level, wage growth leads to inflation. If the rate of growth of wages slows down, we might expect the labour market to be looser; in other words, the demand for labour in relation to the supply of labour is falling. Under normal market conditions, we would expect the wage rate to start to fall. In reality, this translates to the slowing down in the rate of growth of wages.

We then saw how further research into the Phillips curve led to the development of the so-called expectations augmented Phillips curve, captured in the following formula:

$$\Delta p_t = E_t \Delta p_t + 1 + (NR_t - RU_t)$$

This simply says that the change in the inflation rate in a time period (Δp_t) is equal to expected inflation in a time period ($E_t \Delta p_t + 1$) and the natural rate of unemployment minus the national unemployment rate. This helped to explain what seemed to be happening in many Western economies in the 1970s. Households were *anticipating* inflation in the future and basing their behaviour and decision making on those expectations: therefore, the Phillips curve was shifting. This helped to explain how higher inflation might also be experienced at the same time as higher unemployment – so-called 'stagflation'.

There have, however, been a number of economists who have questioned the existence of the Phillips curve. One economist did pose that very question. David Blanchflower has worked at a number of institutions, including the University of Surrey and Warwick University, in the UK. He is a former member of the Bank of England's Monetary Policy Committee and works at Dartmouth College in New Hampshire in the United States.

Remember that the assumption of the Phillips curve is that there is a relationship between the rate of

growth of wages and unemployment. When unemployment rises, the *rate of growth of wages* falls and vice versa. This makes intuitive sense since if there is a large pool of unemployed workers, employed workers will be hesitant to push for higher wage claims for fear of joining the rank of unemployed.

Blanchflower has been quoted as suggesting that as a result of his work, 'the Phillips curve is wrong, it's as fundamental as that'. The following is a summary of his analysis in support of this view. The traditional Phillips curve relates the rate of growth of wages with the level of unemployment. Within this relationship, a higher level of unemployment is associated with a lower level of the rate of growth of wages. Conversely, if unemployment rates were low, the rate of growth of wages would be higher. Such a relationship is given as a macroeconomic one rather than microeconomic, i.e. it holds for the economy as a whole. If the level of unemployment increases, therefore, it suggests there will be excess supply in the labour market. In such cases, the labour market will adjust and the rate of growth of wages will fall to eliminate the excess supply.

(Continued)

Blanchflower and his colleague, Andrew Oswald, spent time researching links between unemployment and wage rates at a *microeconomic level* and found that the relationship between unemployment and wages might be different depending on the region that was being investigated. Their research looked at the *level of pay rather* than the rate of growth of wages. They argued that the *level of pay* was negatively related to the level of unemployment rather than the rate of growth in pay.

According to this argument, a worker in region A, which has a high level of unemployment, would earn lower wages than an equivalent worker in region B with lower levels of unemployment. Blanchard and Oswald's research casts doubt upon the standard explanation, in both regional economics and labour economics, that the wage rate in an area is positively linked to the level of unemployment in an area. In other words, the higher the level of unemployment, the higher the wage level needed to persuade someone to work in that area and vice versa.

Their work would tend to call into question some of the basic 'laws' of economics, particularly those related to something like the

minimum wage. The conventional wisdom might be that if the minimum wage was introduced into an economy, the higher wage levels would be associated with a rise in unemployment. Blanchflower and Oswald's analysis suggests that this may not be the case and that in some areas there might even be a rise in the level of employment associated with a rise in wage levels.

Figure 34.8 highlights the traditional view of the effect of the imposition of a minimum wage set above the market wage level (W_1) on the level of unemployment. The minimum wage causes a fall in the quantity of labour demanded and an increase in the quantity supplied of labour (a movement along the D and S curves for labour). The result is an increase in the amount of unemployment, shown by the distance $Q3 - Q2$.

To understand this, it is important to refer back to the basic model of the labour market shown in Figure 34.9. If the demand for labour, for example, rose as indicated by a shift in the demand for labour curve to the right ($D_L - D_{L1}$), then there would be an excess demand for labour shown by the distance $Q2 - Q1$. Wage rates would rise in response to this shortage of labour and as the shortage is

competed away, more people would end up being employed ($Q3$) at higher wage rates (W_2). A rise in wage rates, therefore, is positively correlated with a fall in unemployment and vice versa.

Blanchflower and Oswald's research was based around millions of observations. They identified the existence of wage curves in 16 countries and since the publication of their research in 1994, a number of other researchers have confirmed their findings. Blanchflower and Oswald suggested that their wage curve would have an elasticity of -0.1 . What this refers to is the responsiveness of rates of pay to changes in unemployment. An elasticity of -0.1 , therefore, would suggest that if we took two regions within an economy, region A and region B, and if unemployment was 2 per cent in region B but 4 per cent in region A, we would expect wages to be around 10 per cent lower in region A than B. This suggests some sort of causal relationship stemming from the level of unemployment in an area feeding through to the level of wages in that area, rather than the other way round.

The suggestion that there is a relationship between the rate of growth of wages (and by implication, inflation)

FIGURE 34.8

The Traditional Explanation of the Effect of a Minimum Wage Set Above the Equilibrium Wage Level

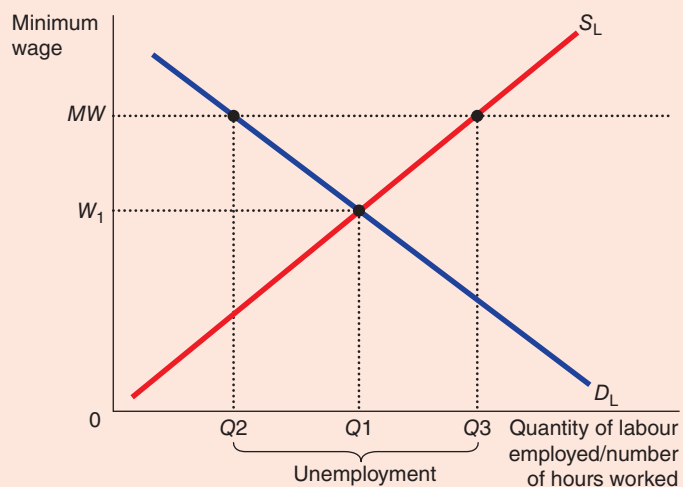
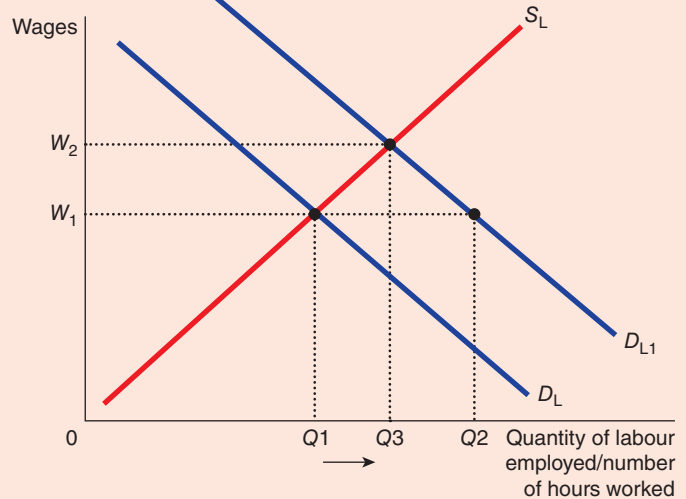


FIGURE 34.9**A Microeconomic Model of the Labour Market**

and unemployment, as we saw in the early part of this chapter, has implications for policymakers. This is based in part on the idea that controlling the rate of growth in wages (and hence inflation) is an important step in controlling unemployment. If, however, Blanchflower and Oswald's research on wage curves is correct, it changes the emphasis of policy making both at a macro and at a micro level. Let us assume that the government are looking at an unemployment rate that they consider too high, and so want to reduce that rate. If they refer to

traditional macroeconomic models of the labour market, they might look at wages as being one way of reducing unemployment in that area – in other words, a fall in wages would help to bring about a reduction in unemployment.

If a wage curve does exist, then attempts to reduce unemployment in this way could backfire, since falling wages would be associated with higher unemployment! In addition, it might be believed that high unemployment in an area might be affected most dramatically by

focusing on high unemployment groups in that area – those with low levels of education, for example. It might be expected that these groups of people will be most likely to take on jobs that are created at lower wages and as such, policies to reduce unemployment could be achieved without a major impact on inflation. Again, the existence of a wage curve would tend to provide an argument against doing this. This is partly because it is the *level* of wages that are important rather than the rate of growth of wages.

INFLATION TARGETING

In many countries of the world today, including the UK and the Euro Area, the framework of monetary policy involves setting targets for inflation to be achieved over the medium run of two years or so, and then adjusting interest rates in order to achieve this objective. As we have pointed out in earlier chapters, setting interest rates is tantamount to setting the money supply if there is a stable and known relationship between the money supply and interest rates. However, the explicit focus on interest rates as the instrument of monetary policy, and medium-run inflation as the target of monetary policy, appears to have worked well for many economies. In the UK the move to the framework of inflation targeting arose as much out of the lessons of experience as out of the logic of economic reasoning.

Targeting

The level of the money supply and the exchange rate can be thought of as intermediate targets of monetary policy. That is to say, they may be targeted in order to achieve certain other, policy targets, such as real GDP growth, employment and inflation. Now, an important principle that we have learnt in

our study of monetary policy is that, in the long run, monetary policy cannot affect the real side of the economy – i.e. real GDP and employment. In the long run, the only final target of monetary policy can be the inflation rate. Therefore, targeting the money supply or the exchange rate as a framework for monetary policy means that we are targeting those variables because it is felt that they will ultimately affect the rate of price inflation. Also, since neither the money supply nor the exchange rate is directly controllable, they cannot be instruments of monetary policy. In other words, the government cannot just set the level of the money supply it wants, it must attempt to bring about that level by open market operations and other means; similarly, it cannot just set the level of the exchange rate it wants, it must try to bring about that level by setting interest rates and by other means. Hence, the money supply and the exchange rate should really be viewed as intermediate targets that the government may aim for because it is believed that they will ultimately affect the final target of inflation. So why not just target the inflation rate? Interest rates (or at least some interest rates) are under the control of central banks, so they are definitely an instrument of monetary policy. And we know that raising interest rates reduces aggregate demand and so dampens inflation. So the government could just raise interest rates whenever inflation rose above a level that was considered desirable – say 2 per cent. This would be a crude form of inflation targeting.

There are two problems with this crude inflation targeting. First, changing (or not changing) interest rates today will not affect inflation today, it will only affect inflation in the future, since it takes time for the policy to have an effect on the economy. Second, there may be other factors that would affect inflation in the future. Suppose, for example, that inflation was at its target of 2 per cent per year. The central bank therefore decides not to change interest rates, since it is achieving its inflation target. However, suppose there was at the same time a wave of very high wage settlements in the economy, above any increases in labour productivity. Over the next six months to a year, this is expected to lead to a rise in price inflation in the economy as workers spend more and as firms pass on some of the wage increases in the form of higher prices. Should the central bank wait for the inflation to arrive, or should it act now by raising interest rates in anticipation of the inflationary pressures coming from higher wages? Clearly, the better policy is to target not today's inflation, which it is already too late to have any effect on, but future inflation. Of course, no one knows with certainty what future inflation is going to be, therefore a policy of inflation targeting generally involves targeting the forecast rate of inflation.

The inflation target in the UK at the time of writing is 2 per cent per year, as measured by the CPI, and to some extent, success in achieving the target is measured by whether the Bank of England's own inflation forecast up to two years ahead falls within 1 per cent of the target, above or below. Of course, it might be argued that the Bank of England is having it easy in the sense that it can judge itself whether or not it has achieved the target. However, an important element of successful inflation targeting is transparency: the Bank of England publishes detailed analyses of its forecasts and these are closely scrutinized. Second, if actual inflation does ever fall outside the range of plus or minus 1 per cent of the target, the Governor of the Bank of England is required to write a formal letter to the Chancellor of the Exchequer, explaining why inflation is out of the target range, what measures are being taken to get it back on course, and when it is expected to be back on target.

As we discussed in Chapter 26, both the Bank of England and the European Central Bank now pursue a policy of inflation targeting, and both target a level of inflation of 2 per cent. The country that pioneered inflation targeting was New Zealand, which adopted this policy in 1990. It was followed in 1991 by Canada, and in 1992 by the UK and Sweden. Like the UK, Sweden also adopted a policy of inflation targeting after it was forced to abandon its exchange rate peg in 1992. Other countries that have adopted inflation targeting in recent years include Norway and Switzerland.

CONCLUSION

This chapter has examined how economists' thinking about inflation and unemployment has evolved over time. We have discussed the ideas of many of the best economists of the 20th century: from the Phillips curve of Phillips, Samuelson and Solow, to the natural-rate hypothesis of Friedman and Phelps, to the rational expectations theory of Lucas, Sargent and Barro. Five of this group have already won Nobel prizes for their work in economics, and more are likely to be so honoured in the years to come. Our discussion

has been based around the UK economy but the principles apply across other economies too – the UK is a convenient case study in the changing interpretation and application of economic theory.

Although the trade-off between inflation and unemployment has generated much intellectual turmoil over the past 40 years, certain principles have developed that today command consensus. Here is how Milton Friedman expressed the relationship between inflation and unemployment in 1968 (in an *American Economic Review* paper entitled 'The Role of Monetary Policy'):

There is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off. The temporary trade-off comes not from inflation per se, but from unanticipated inflation, which generally means, from a rising rate of inflation. The widespread belief that there is a permanent trade-off is a sophisticated version of the confusion between 'high' and 'rising' that we all recognize in simpler forms. A rising rate of inflation may reduce unemployment, a high rate will not.

But how long, you will say, is 'temporary'? ... We can at most venture a personal judgment, based on some examination of the historical evidence that the initial effects of a higher and unanticipated rate of inflation last for something like two to five years.

Today, nearly 40 years later, this statement still summarizes the view of most macroeconomists.

IN THE NEWS

Inflation Targeting

Having a target for inflation has been a mainstay of central bank policy over the last 20 years in many countries. However, the protracted economic difficulties following the financial crisis have raised questions about whether such targets ought to be abandoned.

Should Central Banks Abandon Inflation Targeting?

Ever since the Bank of England was granted independence in 1997, its main policy task has been to maintain inflation at a target level set by the government. Inflation targeting was first introduced in 1992 in the UK. Initially the target rate for inflation was 2.5 per cent but since 2003 has been 2.0 per cent. The average UK inflation rate from 1993–2013 has been 2.1 per cent but in the period from 2008, the average has been 3.2 per cent – well above the 2.0 per cent target. Having a target is meant to send out a signal to everyone that the central bank will act to bring inflation to as close to the target as possible. If inflation persists

above target interest rates may have to be raised to help achieve that target. But, in recent years the idea of raising interest rates has not been top of the Monetary Policy Committee (MPC) agenda; in a period where the economy of the UK has struggled, the idea of raising interest rates would be seen as potentially damaging to the economy.

This has raised the question of whether it is now time to abandon inflation targeting. In July 2013, the Bank of England got a new governor, Mark Carney. In the six months prior to him taking up his position he sparked a debate about whether his tenure would see him follow the inflation targeting which his predecessor, Mervyn King, believed in so strongly. Indeed, Mr Carney, giving

evidence before a Treasury Select Committee in February 2013 noted that under his watch, the Bank of England would follow in the footsteps for the US Federal Reserve in keeping interest rates low to ensure economic recovery was firmly underway and if that meant higher inflation in the process then that would be a price worth paying.

There were also suggestions that the UK Chancellor of the Exchequer would change the targets in the March 2013 budget and ask the Bank of England to focus on the twin targets of employment and inflation. This would allow the Bank of England to set interest rates with the aim of achieving set numerical targets for unemployment regardless of what happens to inflation. There has also

(Continued)

been talk that the Bank of England should adopt a policy of targeting nominal GDP (which effectively has the effect of reducing unemployment). The consequences, however, would be higher inflation (most economists agree with this) which in a time of weak economic growth might be seen as a reasonable trade-off. Question is, would the higher inflation be seen as temporary and once

Mervyn King was the governor of the Bank of England until July 2013.



recovery was well under way and strong, how would people react to the central bank tightening policy once again to bring inflation back under control? Would the credibility of the central bank be destroyed?

Some economists have already put forward the argument that the Bank of England abandoned inflation targeting around 2009 as the consequences for the economy of the financial crisis became more obvious, and it was clear that a period of continued sluggish growth and periodic recessions was in prospect. The fact that inflation has averaged well above the 2.0 per cent target is testament to this, they argue, and note that the explanatory letters from the Governor to the Chancellor of the Exchequer have been 'unconvincing'.

Mervyn King expressed some concern at the idea of abandoning inflation targeting and cited the experience of the 1960s when the UK targeted increasing economic growth which King said was unrealistic and unsustainable in that the price was higher inflation.

Questions

- 1 Explain how monetary policy can be used to target inflation.**
- 2 The average inflation rate in the UK has been just over the current target of 2.0 per cent over the 20-year period from 1993–2013. Should the success of inflation targeting be judged over such a period of time or is the fact that inflation has been much higher than target in the five years from 2009 to 2013 been more indicative of the success of monetary policy?**
- 3 How would a central bank go about setting policies designed to target nominal GDP at a growth rate of (say) 5 per cent?**
- 4 Why would central bank action to boost economic growth be likely to lead to more rapidly accelerating inflation?**
- 5 Given the difficult economic conditions in many countries following the financial crisis, should central banks abandon inflation targeting and pursue a policy of growth?**

SUMMARY

- The Phillips curve describes a negative relationship between inflation and unemployment. By expanding aggregate demand, policymakers can choose a point on the Phillips curve with higher inflation and lower unemployment. By contracting aggregate demand, policymakers can choose a point on the Phillips curve with lower inflation and higher unemployment.
- The trade-off between inflation and unemployment described by the Phillips curve holds only in the short run. In the long run, expected inflation adjusts to changes in actual inflation, and the short-run Phillips curve shifts. As a result, the long-run Phillips curve is vertical at the natural rate of unemployment.
- The short-run Phillips curve also shifts because of shocks to aggregate supply. An adverse supply shock, such as the increase in world oil prices during the 1970s, gives policymakers a less favourable trade-off between inflation and unemployment. That is, after an adverse supply shock, policymakers have to accept a higher rate of inflation for any given rate of unemployment, or a higher rate of unemployment for any given rate of inflation.
- When the central bank contracts growth in the money supply to reduce inflation, it moves the economy along the short-run Phillips curve, which results in temporarily high unemployment. The cost of disinflation depends on how quickly expectations of inflation fall. Some economists argue that a credible commitment to low inflation can reduce the cost of disinflation by inducing a quick adjustment of expectations.

QUESTIONS FOR REVIEW

- 1 Phillips originally looked at the rate of change in money wages rather than inflation but the Phillips curve has come to be associated with the trade-off between inflation and unemployment. What is the relationship between the rate of change of money wages and inflation?
- 2 Explain why, in the short run, there might be a trade-off between inflation and unemployment.
- 3 Draw the short-run trade-off between inflation and unemployment. How might the central bank move the economy from one point on this curve to another?
- 4 Draw the long-run trade-off between inflation and unemployment. Explain how the short-run and long-run trade-offs are related.
- 5 Why do economists place such importance on the role of expectations in macroeconomic policy?
- 6 What's so natural about the natural rate of unemployment? Why might the natural rate of unemployment differ across countries?
- 7 Explain how expected inflation might cause a shift in the short-run Phillips curve.
- 8 Suppose a drought destroys farm crops and drives up the price of food. What is the effect on the short-run trade-off between inflation and unemployment?
- 9 What reasons are put forward for keeping central banks independent in setting monetary policy?
- 10 The central bank decides to reduce inflation. Use the Phillips curve to show the short-run and long-run effects of this policy. How might the short-run costs be reduced?

PROBLEMS AND APPLICATIONS

- 1 Suppose the natural rate of unemployment is 6 per cent. On one graph, draw two Phillips curves that can be used to describe the four situations listed here. Label the point that shows the position of the economy in each case.
 - a. Actual inflation is 5 per cent and expected inflation is 3 per cent.
 - b. Actual inflation is 3 per cent and expected inflation is 5 per cent.
 - c. Actual inflation is 5 per cent and expected inflation is 5 per cent.
 - d. Actual inflation is 3 per cent and expected inflation is 3 per cent.
- 2 Illustrate the effects of the following developments on both the short-run and long-run Phillips curves. Give the economic reasoning underlying your answers:
 - a. a rise in the natural rate of unemployment
 - b. a decline in the price of imported oil
 - c. a rise in government spending
 - d. a decline in expected inflation.
- 3 Suppose that a fall in consumer spending causes a recession.
 - a. Illustrate the changes in the economy using both an aggregate supply/aggregate demand diagram and a Phillips curve diagram. What happens to inflation and unemployment in the short run?
 - b. Now suppose that over time expected inflation changes in the same direction that actual inflation changes. What happens to the position of the short run Phillips curve? After the recession is over, does the economy face a better or worse set of inflation–unemployment combinations?
- 4 Suppose the economy is in a long-run equilibrium.
 - a. Draw the economy's short-run and long-run Phillips curves.
 - b. Suppose a wave of business pessimism reduces aggregate demand. Show the effect of this shock on your diagram from part (a). If the central bank undertakes expansionary monetary policy, can it return the economy to its original inflation rate and original unemployment rate?
 - c. Now suppose the economy is back in long-run equilibrium, and then the price of imported oil rises. Show the effect of this shock with a new diagram like that in part (a). If the central bank undertakes expansionary monetary policy, can it return the economy to its original inflation rate and original unemployment rate? If the central bank undertakes contractionary monetary policy, can it return the economy to its original inflation rate and original unemployment rate? Explain why this situation differs from that in part (b).
- 5 Suppose the central bank believed that the natural rate of unemployment was 6 per cent when the actual natural rate was 5.5 per cent. If the central bank based its policy decisions on its belief, what would happen to the economy?
- 6 Suppose the central bank announced that it would pursue contractionary monetary policy in order to

reduce the inflation rate. Would the following conditions make the ensuing recession more or less severe? Explain.

- a. Wage contracts have short durations.
 - b. There is little confidence in the central bank's determination to reduce inflation.
 - c. Expectations of inflation adjust quickly to actual inflation.
- 7** Some economists believe that the short-run Phillips curve is relatively steep and shifts quickly in response to changes in the economy. Would these economists be more or less likely to favour contractionary policy in order to reduce inflation than economists who had the opposite views?
- 8** Imagine an economy in which all wages are set in three-year contracts. In this world, the central bank announces a disinflationary change in monetary policy to begin immediately. Everyone in the economy

believes the central bank's announcement. Would this disinflation be costless? Why or why not? What might the central bank do to reduce the cost of disinflation?

- 9** Given the unpopularity of inflation, why don't elected leaders always support efforts to reduce inflation? Economists believe that countries can reduce the cost of disinflation by letting their central banks make decisions about monetary policy without interference from politicians. Why might this be so?
- 10** Suppose policymakers accept the theory of the short run Phillips curve and the natural-rate hypothesis and want to keep unemployment close to its natural rate. Unfortunately, because the natural rate of unemployment can change over time, they aren't certain about the value of the natural rate. What macroeconomic variables do you think they should look at when conducting monetary policy?

35

SUPPLY-SIDE POLICIES

The focus of policy in the post-war era was firmly on managing aggregate demand. In the late 1960s and into the 1970s demand management came to be questioned and some economists began looking at the capacity of the economy – aggregate supply. In particular, economists began looking at the effects of (often) frequent changes in taxes which was a feature of demand management on the supply-side of the economy.

We have seen that in the long run, the aggregate supply curve is vertical but in the short run it slopes upwards from left to right. Shifting the AS curve to the right would increase the capacity of the economy and increase national income (and thus lower unemployment) but reduce pressure on prices. A focus on the supply-side of the economy could, therefore, help to bring about sustainable economic growth; higher growth, lower unemployment and stable or even lower inflation. Supply-side policies, therefore, focused on ways of influencing the factors which affect aggregate supply. If aggregate supply represents the quantity of goods and services that firms choose to produce at different price levels then looking at what influences firms' behaviour is an important starting point.

SHIFTS IN THE AGGREGATE SUPPLY CURVE

There is a range of factors which can cause the aggregate supply curve to shift as we saw in Chapter 32. We summarized these as changes in labour, physical or human capital, the availability of natural resources, changes in technology and expectations of the price level. This summary provides us with the bare bones on which to explore in more detail how shifts in the aggregate supply curve might arise.

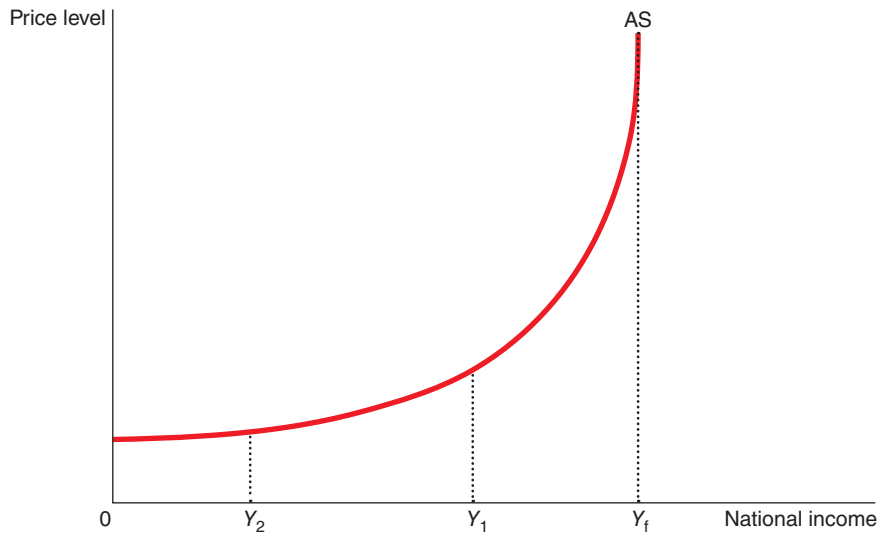
So far we have looked at the AS curve in the short run as being upwards sloping and in the long run as vertical. The vertical long-run AS curve makes intuitive sense if you think about AS in terms of the capacity of the economy. Imagine that every individual in an economy who wanted to work at the going wage rate could find work (i.e. there was full employment), that every firm was producing at its maximum capacity and there were no machines lying idle then the economy would be producing at its maximum capacity which we will refer to as full employment output (Y_f).

We know that in the short run the AS curve can shift causing temporary deviations from this full employment output but eventually, once changes in prices, wages and misperceptions work through the economy, output returns to its long-run equilibrium. Changes in AD can cause fluctuations in economic activity and governments can use fiscal and monetary policy to manage demand to try and stabilize the fluctuations. However, if policies are put in place to boost AD, for example, the effects might depend on the extent of the output gap that exists and the flexibility of the economy to respond to the increase in AD. The extent to which unemployment is reduced and how far the price level changes will be affected by the flexibility of the economy.

Figure 35.1 shows a different shaped AS curve to those we have looked at so far. At the full employment level of output, the AS curve is vertical but between Y_1 and Y_f , the AS curve becomes increasingly steep, whereas between Y_2 and Y_1 , the AS curve is almost horizontal. This is referred to as the new Keynesian aggregate supply curve.

FIGURE 35.1**The New Keynesian Aggregate Supply Curve**

The new Keynesian aggregate supply curve is vertical at full employment output Y_f . At low levels of national income, the AS curve is virtually horizontal reflecting the fact that there is considerable spare capacity in the economy but, as national income rises, the ability of firms to expand output becomes more difficult and so the AS curve becomes gradually steeper until it becomes vertical at full employment output.



Between the origin and Y_2 , there is a considerable amount of spare capacity in the economy but that spare capacity begins to diminish the closer to full employment output at Y_f . This has implications for the effect on national income and the price level if aggregate demand shifts as shown in Figure 35.2. If the economy starts out at Y_1 with the price level P_1 , there is considerable spare capacity in the economy and the output gap is large. If fiscal and/or monetary policy shifts the AD curve to the right, the increase in national income rises by a relatively large amount from Y_1 to Y_2 . The increase in the price level is only small, suggesting that the Phillips curve is relatively shallow and the trade-off between the reduction in unemployment and the rise in the price level is favourable. If policy continues to push AD to the right, the trade-off changes and the increase in national income begins to get smaller the nearer to full employment output whilst the rise in the price level increase at a faster rate.

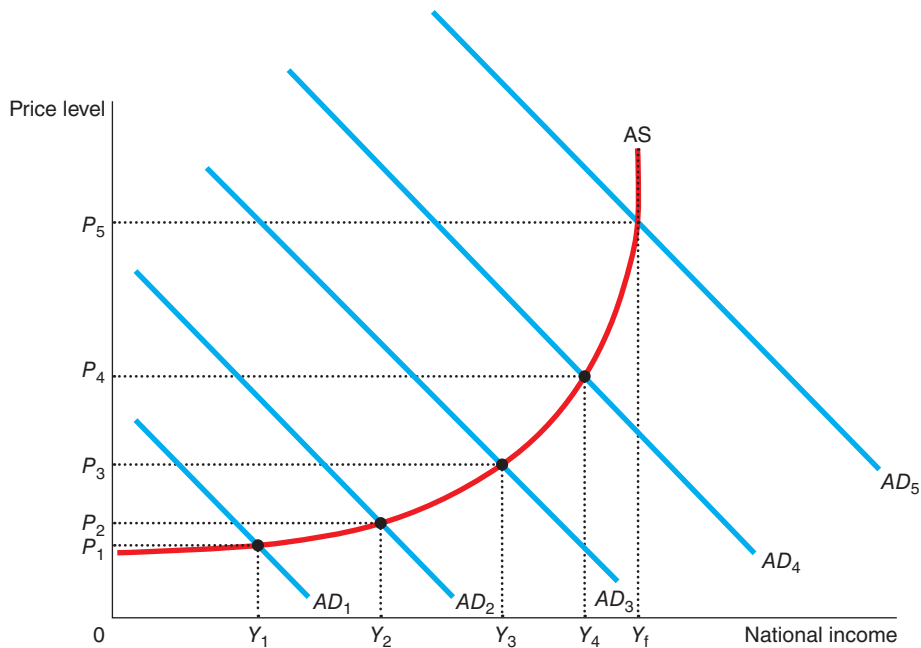
Why does the trade-off become less favourable over time? The reason is that as aggregate demand increases, the pressure on society's scarce resources becomes greater. If AD increased from AD_3 to AD_4 , for example, firms would be looking to expand output to meet the increased demand and so would look to buy more factor inputs. As the demand for these factor inputs rises, the price of them increases. Certain types of labour might become scarce more quickly than others and wage rates in these industries might rise sharply as a result. The demand for skilled labour in the construction industry, for example, might rise but the supply of people with these skills tends to be inelastic in the short run. It takes time to train skilled bricklayers, plumbers, plasterers and electricians. Those who have these skills will find them in demand and they can sell their labour for higher wages as a result. Firms that pay the higher wages see their costs increase but with demand rising they feel they can pass on these higher costs to the consumer in the form of higher prices and so inflation begins to accelerate. The closer the economy comes to full employment output the greater will be the production bottlenecks that arise as resources become scarcer so the effect on prices will be greater but the ability of the economy to grow further will decline.

The supply-side focus arises because of the difficulty of policymakers knowing what the size of the output gap is at any particular moment in time; that is, where the economy is on the AS curve. The output gap can be calculated in two ways. The first is to subtract actual output from potential output where potential output is based on the assumption that markets are perfectly competitive. The second is to subtract

actual output from natural output where natural output assumes imperfect competition and rigidities in prices and wages. We have seen how monetary and fiscal policy operate side-by-side and even if governments are not deliberately trying to influence aggregate demand, changes in tax rates, allowances and government spending decisions will have effects on the level of aggregate demand, whilst changes in monetary policy affect investment decisions and the exchange rate, and, as a result, net exports. If the net effect of monetary and fiscal policy is to boost aggregate demand, the size of the output gap will have an effect on the resulting changes in the price level and national income.

FIGURE 35.2**The Aggregate Supply Curve and Production Bottlenecks**

Increases in AD when there is considerable spare capacity in the economy can lead to relatively large increases in national income with only relatively small rises in the price level. If AD keeps increasing resource constraints will eventually mean that national income will only rise by relatively small amounts but the pressure on resources will cause the price level to rise more quickly.



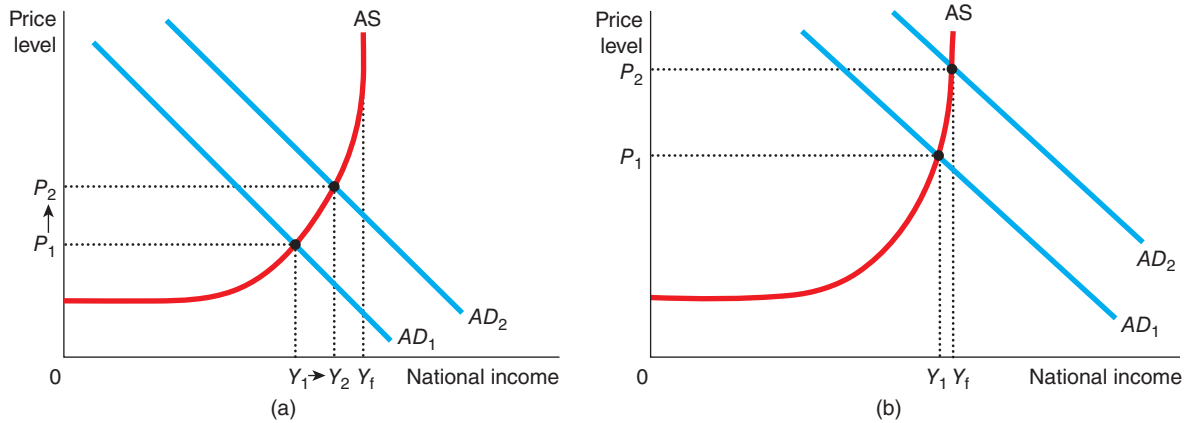
This is illustrated in Figure 35.3. Assume the economy in panel (a) is in equilibrium at Y_1 . The output gap is given by the distance between full employment output and actual output $Y_f - Y_1$. An increase in AD to AD_2 would lead to a rise in national income to Y_2 and a rise in the price level to P_2 . The output gap narrows but the economy is beginning to experience some bottlenecks in production and as a result the increase in national income has come at a price of accelerating inflation. Any further increases in AD would mean the economy is in danger of overheating – where supply-side constraints mean that the pressure on prices increases.

In panel (b), the initial equilibrium is given by Y_1 but the output gap, again given by $Y_f - Y_1$, is much smaller. The economy is already nearing the vertical section of the AS curve and an increase in AD means that the increase in national income is very small but the effect on the price level is much stronger. The economy is overheating and does not have the capacity to be able to cope with the rise in AD and so the main effect feeds through to prices.

The size of the output gap is simple in practice but much more difficult in reality to estimate. The models used vary and the size of the various parameters that modellers input into the equations can lead to widely differing results. If policymakers base decisions on an estimated output gap that is incorrect the effects of policy can be different to those anticipated. It is for this reason that supply-siders argue that a focus on increasing the productive capacity of the economy needs to be an important element of the policy mix.

FIGURE 35.3**The Effect of the Size of the Output Gap**

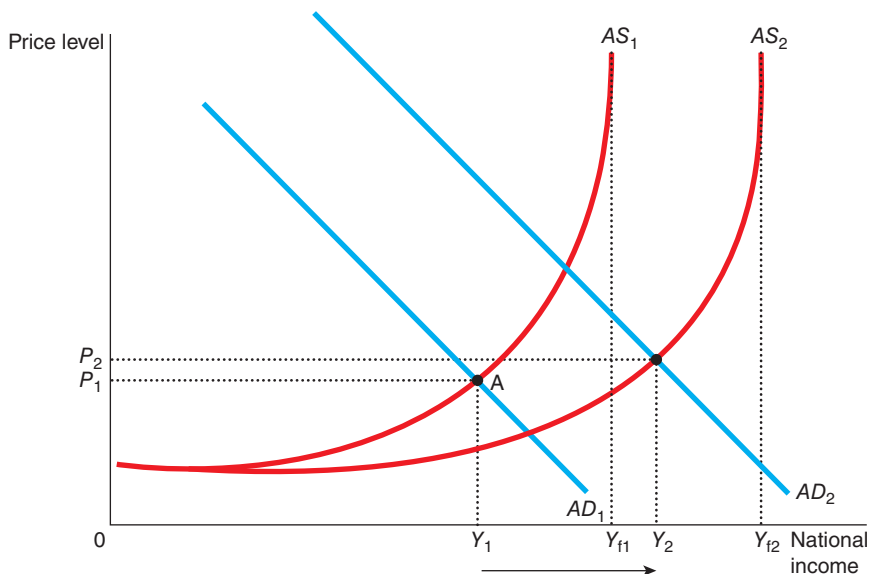
Panel (a) shows the effect on national income and the price level of a rise in aggregate demand where the output gap is relatively large. In contrast, if the economy is in an initial equilibrium when the output gap is small as in panel (b), then a rise in AD will have a much larger effect on the price level.



Supply-side economists suggest that if policy is focused on increasing the capacity of the economy then in the long run it is possible to increase national income but keep the price level more stable. This is illustrated in Figure 35.4. Assume society faces an initial AS curve, AS_1 , and AD curve, AD_1 . The initial equilibrium is at point A with a price level of P_1 and national income at Y_1 . An increase in AD to AD_2 would, given the initial AS curve, see the economy begin to experience more significant production bottlenecks

FIGURE 35.4**Sustained Economic Growth**

If supply-side policies are successful in shifting the AS curve from AS_1 to AS_2 , an increase in aggregate demand from AD_1 to AD_2 results in a relatively large increase in national income from Y_1 to Y_2 but with only a modest increase in the price level from P_1 to P_2 . The new AS curve is associated with an increased capacity of the economy shown by the increase in full employment output from Y_{f1} to Y_{f2} .



and as a result the price level would start to accelerate. If supply-side policies succeeded in shifting the AS curve to the right to AS_2 , this would give a new full employment output capacity at Y_{12} compared to Y_{11} . The increase in aggregate demand to AD_2 would now result in a rise in national income to Y_2 from Y_1 but the increase in the price level would be relatively small rising from P_1 to P_2 . Many economists would argue that a little inflation in an economy is a good thing as it acts as an incentive to firms to produce and expand. By shifting AS to the right, the economy could increase standards of living and reduce unemployment but keep prices stable (where price stability is defined as having an inflation rate which was deemed acceptable – i.e. the target inflation rate). This would represent sustained economic growth.

SELF TEST What are the factors which can cause a shift in the AS curve to the right? What is meant by ‘production bottlenecks’?

TYPES OF SUPPLY-SIDE POLICIES

Supply-side economists have a fundamental belief in the power of markets and the private sector to allocate resources efficiently – at least more efficiently than might be the case with public sector provision. Policies, therefore, focus on the extent to which the working of markets can be improved to ensure more efficient allocation of resources. These policies are classed as **market-orientated supply-side policies**. If governments are going to intervene in the economy then supply-siders believe that the focus should be on investing in resources that improve the working of the economy such as developing infrastructure, improving research and development and investing in education. These policies are referred to as **interventionist policies**. We will now look at the policies which come under these broad classifications in more detail.

market-orientated supply-side policies policies designed to free up markets to improve resource allocation through more effective price signals

interventionist supply-side policies policies focused on improving the working of markets through investing in infrastructure, education and research and development

Market Orientated Supply-Side Policies

The intention of market orientated supply-side policies is to help markets work more effectively by improving the effectiveness of price signals so that resources are allocated more efficiently. Underpinning these policies is a philosophical belief that markets are the most effective way of allocating resources. The use of Keynesian demand management policies had led to government becoming more and more involved in the economy, with a number of industries or firms run by the state on behalf of the people as a whole. Many governments increasingly built up the welfare state to support those in need and, of course, all this government spending meant taxes had to be raised to pay for the increase in the services provided by the state.

Supply-side policies came to the fore in the 1980s when governments in the developed world were trying to cope with stagflation. This section is not intended to be a history lesson but there will be references to policies put in place by the governments in the United States and the UK to highlight key issues arising in supply-side policies. A core focus of what was referred to as the ‘supply-side revolution’ was on rolling back the influence of the state in the economy, cutting government spending, reforming the tax and benefit system to improve incentives, freeing up the labour market to make it more flexible and encouraging a more entrepreneurial culture in society. Many of the policies implemented were designed to improve incentives, which, as we know, is one of the *Ten Principles of Economics*.

Reform of Tax and Welfare Policy We have seen that taxes have an effect on incentives and can lead to distortions in market outcomes. Welfare payments can be viewed as negative taxes giving people support when they fall on hard times. Few would disagree with the necessity of taxing firms, products and individuals to generate tax revenue for government and also to influence behaviour in desired directions such as reducing smoking or encouraging reductions in waste. If we acknowledge the fact that taxes – positive and negative – distort market outcomes, then the debate shifts to the level of taxation.

In many countries in Europe and in the US, income and corporate taxes have been relatively high in the 1970s and 1980s. In the UK, for example, the basic rate of income tax was 33 per cent in 1978, there were also other higher rates associated with higher levels of income, with the top tax rate being 98 per cent on unearned income and 83 per cent on earned income. Corporation tax in the UK in 1978 was 53 per cent for larger companies and 42 per cent for small companies. In the US, income tax rates ranged from 14 per cent to 70 per cent and corporate tax rates were near 50 per cent. Corporate taxes in Germany in 1981 were around 56 per cent, in Norway just over 50 per cent, 48 per cent in the Netherlands and in Sweden almost 58 per cent. Table 35.1 shows the top rates of income tax in a selection of nine European countries for 1979, 1995 and 2012.

TABLE 35.1**Top Rates of Income Tax in Nine European Countries, 1979–2012**

| Country | Top Income tax rate 1979 (%) | Top Income tax rate 1995 (%) | Top Income tax rate 2012 (%) |
|-------------|------------------------------|------------------------------|------------------------------|
| Belgium | 76.3 | 55.0 | 50.0 |
| Denmark | 66.0 | 63.5 | 55.0 |
| Germany | 56.0 | 53.0 | 45.0 |
| Ireland | 60.0 | 48.0 | 48.0 |
| Netherlands | 72.0 | 60.0 | 52.0 |
| Norway | 75.4 | 13.7 | 47.8 |
| Portugal | 80.0 | 40.0 | 46.5 |
| Sweden | 86.5 | 30.0 | 56.6 |
| UK | 83.0 | 40.0 | 50.0 |

Supply-siders argued that high rates of income and corporate taxes were acting as a disincentive to individuals and to businesses. We need here to remind ourselves of the distinction between average and marginal tax rates. Remember that marginal tax rates refer to the amount of tax taken on every extra euro of earnings. In the UK, some people faced a marginal tax rate of 83 per cent – the government would leave the individual with just 17 pence from every extra pound earned. Supply-siders argued that such high rates of tax acted as a significant disincentive to work and to be entrepreneurial if the rewards from hard work were subject to such rates.

Cutting both income and corporate taxes, therefore, could lead to the incentive for people to take work and work harder (more hours) if the rewards were greater. Income tax rates, it was argued, altered the trade-off between work and leisure. Higher tax rates reduce the opportunity cost of leisure and so people will opt to reduce work and take more leisure. If tax rates are lowered and people get to keep more of the income they earn, the effect is like an increase in wages. The substitution effect is reinforced by the income effect and labour supply increases.

High corporate and income taxes also act as an incentive for firms and individuals to engage in tax avoidance (which is legal) or tax evasion (which is illegal). Higher tax rates, it is argued, do encourage firms and individuals to find ways in which their tax liability can be reduced, which may include putting money into offshore ‘tax havens’, firms spending money on tax-deductible assets which may not have any major impact on productivity (like investing in plush office suites, buying luxury company cars, locating conferences at more luxurious venues and so on), or on channelling income into offshore companies which then return the money in the form of ‘loans’ on which there is no income tax. Such schemes are not illegal but some would claim that they are morally wrong. Tax evasion is illegal and refers to attempts by firms and individuals to avoid declaring income which is subject to taxation. This might be in the form of traders

accepting cash for work carried out which is then not declared as income, through to more sophisticated money laundering schemes. Estimates in the UK, for example, put the cost of tax evasion at around £14 billion a year and some research by the Tax Justice Network in 2011 put the cost of tax avoidance across 145 countries at \$3.1 trillion (€2.39 trillion or £2.03 trillion).

Reducing tax rates can change incentives to work and also incentives to avoid or evade paying taxes. Cutting tax rates can, therefore, mean that tax revenues actually increase. An example serves to illustrate. Assume that the top rate of income tax in a country is 80 per cent and that this rate applies to the whole of an individual's income (in reality most income tax regimes in country's are progressive and stepped so that those on lower incomes pay a lower rate compared to those on higher rates). An individual earning €500,000 would have to pay €400,000 in tax and gets to keep the remaining €100,000. If the government reduced the tax rate to 60 per cent, the individual faces the incentive that they get to keep more of every €1 of earnings which acts as a powerful incentive to work harder. The individual works harder in the next year and earns €750,000, on which her pay tax of €450,000 and gets to keep €300,000. The individual has more of an incentive to work harder because she gets to keep more of her earnings and the government generates more in overall tax revenue.

Income and corporate tax rates fell in many countries across Europe in the 1980s and into the 1990s. In addition to reducing actual tax rates, governments can also change tax allowances which relate to the amount of income which can be earned before having to pay tax. In the UK, for example, the Coalition Government, which took office in 2010, changed tax allowances so that workers will not pay tax on the first £10,000 of income compared to an allowance of around £6,500 which was in existence when it took office. Changing tax rates not only acts as an incentive to work harder and be more entrepreneurial, it is argued, it also has an effect on those who are looking for work. If tax rates are high or tax allowances are low, it can be the case that the incentive for people to go into work and come off benefits is also low. Assume, for example, an individual was receiving €10,000 a year in various benefits but pays no income tax. A job becomes available which pays €12,000. If the worker takes the job but then becomes liable for tax at a rate of (say) 25 per cent on all his earnings, then he will receive an after-tax income of €9,000. There is no incentive for the person to take the job, to come off benefits and pay income tax. If the tax rate were reduced to (say) 10 per cent, the person would receive an after tax income €10,800 and so there would be more of an incentive to take the job and come off benefits. Not only would the government be better off by not having to pay this individual benefits, it would be receiving €1,200 in tax revenue – a net benefit to the government of €11,200. In addition, by getting a job the individual is now not subject to the same issues and problems we outlined as the costs of unemployment to individuals in Chapter 23, which reduces the social externalities associated with unemployment.

This approach is also linked with reform of the welfare system to provide incentives to individuals to get off benefits and into work. By making access to benefits subject to more rigorous tests and/or reducing welfare payments, individuals are forced to reassess their position, not become so reliant on the welfare state and look to find ways to support themselves. This means they might have to be more flexible in taking jobs which they may otherwise have shunned because they did not match to their expectations, wage demands or qualifications.

The combined effects of these policies mean that the labour market is more efficient with fewer people being voluntarily unemployed, with those in work prepared to work harder and be more productive and for enterprise and initiative to be rewarded.

The analysis so far is fairly intuitive but the reality depends on a number of factors not least the elasticity of supply of labour. If the elasticity of supply of labour is relatively low then any increase in wages through tax cuts or changes in allowances will only have a limited effect on the increase in the supply of labour (often measured in labour hours). If the elasticity of supply of labour was 0.1, for example, a rise in the wage rate of 5 per cent would only lead to an increase in labour supply of 0.5 per cent. The long run elasticity of supply of labour, however, might be larger than those in the short run because workers have time to adjust their behaviour. Research by Nobel Prize-winning economist, Edward Prescott, in 2002 suggested that there was a considerable difference between the elasticities of labour supply in France and the US which were largely due to differences in tax systems between the two countries.

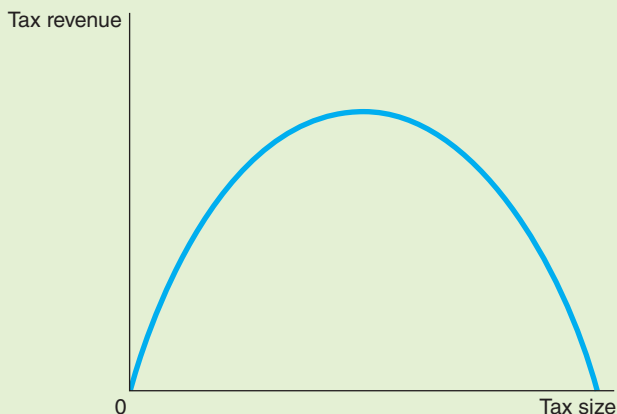
CASE STUDY The Laffer Curve and Supply-Side Economics

One day in 1974, rumour has it that the American economist, Arthur Laffer, sat in a Washington restaurant with some prominent journalists and politicians. He took out a napkin and drew a figure on it to show how tax rates affect tax revenue. The illustration looked like that shown in Figure 35.5. Laffer then suggested that the United States was on the downwards sloping side of this curve. Tax rates were so high, he argued, that reducing them would actually raise tax revenue.

FIGURE 35.5

The Laffer Curve

The Laffer curve shows how tax revenue changes with the tax size. When the tax size is 0, tax revenue is 0. When the tax size is 100%, tax revenue is 0. Tax revenue varies with tax size in between these extremes. The curve shows that reducing tax size could lead to an increase in tax revenue.



Most economists were sceptical of Laffer's suggestion. The idea that a cut in tax rates could raise tax revenue was correct as a matter of economic theory, but there was more doubt about whether it would do so in practice. There was little evidence for Laffer's view that tax rates – in the United States or elsewhere – had in fact reached such extreme levels.

Nevertheless, the thinking underlying the **Laffer curve** (as it became known) became very influential in policy circles during the 1980s, particularly in the USA during the years of President Ronald Reagan's administration and in the UK during Prime Minister Margaret Thatcher's government. Tax rates – particularly income tax rates – were cut aggressively in both countries during the 1980s.

Laffer curve the relationship between tax rates and tax revenue

In the UK, for example, under Prime Minister Thatcher the top marginal rate of income tax was cut from 83 per cent to 60 per cent in 1980 and then again to 40 per cent in 1988. Economists have, however, found it hard to trace any strong incentive effects of these tax cuts leading to increases in total tax revenue as the Laffer curve would suggest. A study by the UK Institute for Fiscal Studies (IFS), for example, concluded that at most about 3 per cent of the increase in tax revenue between 1980 and 1986 could be attributed to the 1980 income tax cut.

In the USA, President Reagan also cut taxes aggressively, but the result was less tax revenue, not more. Revenue from personal income taxes in the United States (per person, adjusted for inflation) fell by 9 per cent from 1980 to 1984, even though average income (per person, adjusted for inflation) grew by 4 per cent over this period. The tax cut, together with policymakers' unwillingness to restrain spending, began a long period during which the US government spent more than it collected in taxes. Throughout Reagan's two terms in office, and for many years thereafter, the US government ran large budget deficits.

Yet Laffer's argument is not completely without merit. Although an overall cut in tax rates normally reduces revenue, some taxpayers at some times may be on the wrong side of the Laffer curve. The idea that cutting taxes can raise revenue may be correct if applied to those taxpayers facing the highest tax rates, but most people face lower marginal rates. Where the *typical* worker is on the top end of the Laffer curve, it may be more appropriate. In Sweden in the early 1980s, for instance, the typical worker faced a



Arthur Laffer – the man behind the Laffer curve.

marginal tax rate of about 80 per cent. Such a high tax rate provides a substantial disincentive to work. Studies have suggested that Sweden would indeed have raised more tax revenue if it had lowered its tax rates.

Policymakers disagree about these issues in part because they disagree about the size of the relevant elasticities. The more elastic that supply and demand are in any market, the more taxes in that market distort behaviour, and the more likely it is that a tax cut will raise tax revenue. There is no debate, however, about the general lesson: how much revenue the government gains or loses from a tax change cannot be computed just by looking at tax rates. It also depends on how the tax change affects people's behaviour.

SELF TEST Consider the respective size of the income and substitution effects on the supply of labour curve as a result of a cut in higher rate taxes. ● Under what circumstances would a cut in tax rates actually lead to a backwards bending supply curve of labour?

Flexible Labour Markets In addition to changing the tax and benefits system, supply-siders argue for more flexibility in labour markets. There are different aspects to this policy which include reducing the power of trade unions and improving market signals to both employers and workers. We saw in Chapter 17 how trade unions can raise wages above market rates. The more powerful are trades unions the higher they can push wages above market rates. In the 1970s and into the 1980s, the power of trade unions in the UK was substantial; days lost to strikes averaged 12.8 million per year in the 1970s. The UK government argued that trade unions exercised too much power and passed a series of new laws which reduced their power considerably. The result has been that trade unions had to adapt and evolve and the number of people belonging to trade unions and the number of days lost to strikes fell sharply in the latter part of the 1980s and beyond. Many more people now negotiate pay levels with employers individually or at local level rather than through unions and supply-siders would argue that wage rates more accurately reflect market conditions.

More flexible labour markets also relate to the ease with which employers can respond to market conditions by adjusting the labour force. This is sometimes referred to as the ability to 'hire and fire'. The more flexible the labour market, the fewer rigidities that would exist.

Flexible labour markets might include the use of short-term contracts with workers, having more flexible working arrangement so that workers with the right skills are attracted to work, and being able to shed workers more easily when the economy is contracting. Critics argue that flexible labour markets can put the interests of employers before those of workers and a balance does have to be struck between the rights of both parties.

Labour markets can also be made more flexible by improving market information for both employers and employees such that those people looking for work can be matched with those seeking employment. Many governments provide so-called 'job centres' through which employers can advertise jobs and those looking for work can attend to get advice, be put in touch with prospective employers and given help with improving their CVs, interview skills, job application techniques and so on.

Reducing Government Spending The more money government spends the more it has to raise in tax revenue and also to borrow. We have seen in Chapter 33 how any fiscal expansion can be moderated by the crowding out effect. It follows that if governments were able to reduce spending and borrow less the opposite effect would occur – sometimes referred to as 'crowding-in'. Cutting back the amount that governments borrow has been a feature of the post-financial crisis austerity programmes which many countries have had to implement. If borrowing can be scaled back, interest rates will fall and this has an effect on firms' decisions to invest.

In addition to the crowding-in argument, there is a fundamental belief among supply-siders that government spending is not likely to be as efficient as the equivalent spending by the private sector. If government, therefore, steps back from its involvement in the economy and leaves resource allocation more to the private sector, the funds will be used more productively and more efficiently which will improve the productive capacity of the economy.

In practice, many governments find it extremely difficult to cut back spending. The effects on government services and those who use it can be considerable and the sort of people who are most affected are often the poor and the most vulnerable in society such as the elderly, the mentally ill, children and those who require social and/or health care. The political decisions which have to be made when government spending is cut back are extremely difficult – few governments can be so sure of the mandate they receive from the electorate to make the sorts of decisions which would lead to a substantial fall in government spending when so many people are likely to be affected in an adverse way. Closing hospitals, schools, social care programmes, homes for the elderly and mentally ill might save considerable sums of money but represent electoral suicide.

Privatization and Deregulation One of the major changes in the economy of the UK in the last 30 years has been **privatization**, the move to transfer public ownership of assets to the private sector and the increase in the role of private sector firms in providing government services. Privatization has also been a feature of other European countries including Austria, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands and Spain.

privatization the transfer of public ownership of assets to the private sector

In the post-war era in the UK, many industries were state owned – coal, the railways, the utilities (water, gas and electricity), steel, and telecommunications being key examples and in addition a number of firms were taken into state ownership including some in the motor, aviation and engineering industries. During the 1980s and 1990s, many of these industries and firms were sold and became private sector entities. In addition other services such as refuse collection, cleaning and catering services were opened up to provision by private sector organizations.

Once again, the philosophical underpinning of privatization in all its forms is the belief that the private sector, driven by providing high quality services and the profit motive, are more likely to provide the services required by the public at lower cost, at higher productivity levels and more efficiently than if provision was in the public sector. It was also argued that transferring assets to the private sector would help increase competition with the resulting benefits of lower prices and increased choice.

It is the subject of much debate as to the extent to which the benefits which were at the heart of the justification for privatization have been realized. The privatization of natural monopolies such as water, gas and electricity has created a complex system which many consumers find difficulty in understanding and despite regulatory bodies being set up to monitor the activities of these now private monopolies,

there is some suspicion that the existence of monopoly power has reduced consumer surplus and led to deadweight losses. There are some studies that have suggested that privatization has delivered better services, improved products, lower prices and more choice, more efficiently for the public.

In addition to privatization programmes, many governments looked to pass legislation which cut back on the regulations that were argued to impede the efficient working of financial and goods markets. Deregulation has succeeded in removing some of these imperfections in the market. For example, in transport, local bus routes have been deregulated to prevent local monopolies which have resulted in more choice and lower prices for consumers. In financial markets, there has been an explosion of new products and much greater freedom for individuals and firms to be able to access credit. Whilst this did have some impact on the global economy in the 1990s and into the 2000s, many have argued that deregulation went too far and that appropriate checks and balances were not in place to prevent the financial crisis in 2007–2008.

SELF TEST Why do policies to improve the flexibility of labour markets have to balance the rights of employers and employees?

Interventionist Supply-Side Policies

In Chapter 10 we noted that the creation of knowledge is a public good and that firms free ride on knowledge created by others and as a result devote too few resources to creating new knowledge. In addition, the investment in infrastructure such as communications networks, roads, rail, ports and airports is vital to the effectiveness of an economy but the sort of sums needed to invest in this sort of infrastructure and the risk involved is often extremely high and the exploitation of the assets for profit by private firms not always easy, obvious or in some cases desirable. For example, the development of new motorway systems raises concerns among local people who are affected by the negative externalities and of environmentalists. The planning process can often take many years and be extremely expensive in addition to the actual cost of construction. The result is that governments often take the lead in the provision of such infrastructure, albeit using private firms as noted above. Interventionist supply-side policies refer to the ways in which governments seek to encourage investment in education, training, research and development and infrastructure and location of firms in order to help the economy work more effectively and efficiently.

Infrastructure Investment Governments will often take the lead in providing funds for infrastructure projects which have been identified as having long-term beneficial effects on the economy and will help to improve capacity. The sort of projects that governments support are invariably identified as being important in helping businesses to be more efficient, cutting costs and waste and improving supply chains. Improvements to roads, rail, ports and airports help firms to distribute goods more efficiently; improvements to communications networks help to make service industries more efficient and developments to energy and water supplies through the development of new power stations or improvements to supply networks can also help firms secure their energy needs and can help reduce cost if these networks operate more efficiently.

Investment in Education and Training The school and higher education system in an economy is a source of improvements in the quality of future human capital. Whilst many would not see education as purely a means of producing future workers, the sort of skills which school and university leavers have influence their ability to secure work and their productivity levels. As we have seen, productivity is a key factor in improving economic growth and standards of living.

The sort of education system that a country has will, therefore, be important and a number of countries have invested in trying to improve standards. The difficulty comes in defining what appropriate standards are for the individuals and for wider society. Some countries place a heavy emphasis on improving maths and science skills because these are crucial in helping to create new knowledge and innovation in industry.

Governments might provide grants or allowances to help boost interest in maths, science and engineering degrees; maths and physics graduates may be given incentives to put their skills to use in teaching in schools where there tends to be a paucity of highly qualified individuals in these subject areas.

Governments may also help to support training; firms may be reluctant to invest too many resources into training workers because of the lack of control they have over what workers do with the training they receive. On the one hand, a better trained workforce is beneficial to a firm because it makes workers more productive. It also makes workers more marketable and there is a risk that the worker moves on to a better job and another firm will benefit from the investment in training made by the initial firm. By providing help and support for training, governments may help to overcome this free-rider problem.

Research and Development R&D is essential to the long-term improvement of knowledge creation, the development of new products and new processes. R&D is, however, expensive and, as noted, unless the results can be protected in some way, the knowledge created becomes a public good. This is one of the reasons why many governments invest in supporting and funding research and development. Research bodies can be set up which receive requests for funding from firms and from higher education institutions, sometimes in partnership, who then decide on the allocation of funds based on the perceived value and importance of the research being proposed. Governments may also provide tax credits, tax relief or grants to help support R&D in smaller firms.

Regional or Industrial Policies In a number of European countries, the economy is now evenly balanced – there are regions which are poorer than others. Firms will often gravitate to locate where the markets are biggest or where there are natural advantages such as ports, good infrastructure links and so on. Some regions are poor because they used to be the centre of industries that have declined. The decline of industries such as shipbuilding, iron and steel, the motor industry and the coal industry, for example, has not been matched by new industries springing up to absorb the jobs lost and as a result a negative multiplier process can take hold which means the region may stay economically undeveloped for long periods. Governments may seek to reverse such economic imbalance by locating some of its own activities in these regions, providing investment grants, premises at reduced rents, employment subsidies and other measures to encourage firms to locate in these areas and develop jobs. The intention is that in encouraging investment in these areas a regional multiplier effect will take hold.

SELF TEST Why is investing in the right sort of education and training considered important to the success of interventionist policies?

CONCLUSION

Supply-side policies have an intuitive feel to them; one of the main problems, however, is that shifting the aggregate supply curve to the right is a long-term process. Investing in education, training and research and development, for example, can boost the quality of human capital but the benefits to the economy may not be felt for many years. Market-orientated policies may sound impressive and laudable but the evidence to support the effectiveness of such policies is not overwhelming. It is perhaps safe to say that policies to improve the efficiency of markets and the capacity of the economy have to be carried out hand-in-hand with other policies and not be seen as being an either/or policy option to fiscal and monetary policy. Indeed, the differences between fiscal, monetary and supply-side policies are sometimes difficult to disentangle. If a government announces it is investing billions in a high-speed rail network does this represent a fiscal boost or is it purely focused on improving the supply-side of the economy?

IN THE NEWS



Supply-Side Policies in an Economic Slowdown

The economic problems facing many countries in the wake of the financial crisis have presented new challenges to policymakers. After several years of attempting to reverse economic decline through fiscal and monetary policy, attention is once again turning to the supply-side option.

Time to Look at the Supply-side Again?

Consider the scenario facing governments in a number of European countries in 2013. Economic growth is barely above zero, inflation remains at levels above target, unemployment in some countries is extremely high but in others, whilst having risen, has not reached the sort of levels predicted or implied by the sluggish GDP growth rates. Interest rates are at historically low levels but concerns over the size of government debt mean that austerity measures have to be pursued if confidence in the financial markets in the government and its policies is to be maintained. This means that the fiscal options open to governments are very limited; if anything, taxes have to be increased and government spending cut back further if targets to reduce debt are to be realized. With the further options for monetary and fiscal policy looking exhausted do supply-side options offer any hope? Why would supply-side policies fare any better? One suggestion is that the problems facing the UK government in 2013 is that there are major structural deficiencies in the economy. The fact that inflation in the UK had been way above target (an average of 3.2 per cent for the five-year period prior to 2013 compared to a target level of 2.0 per cent) would suggest that demand deficiency is not the problem. The fact that unemployment had not risen to the sort of levels predicted after the financial crisis, accompanied by increases in employment rates, further hint at

the problems facing the UK not being with a lack of demand.

The stagnant economic growth in the UK and a number of other European countries would suggest that solutions lie with further monetary policy options such as more quantitative easing and/or, as had been advocated by the UK Labour party in opposition, that it was pertinent to borrow in the short-term to stimulate growth by spending on projects to kick-start the economy. In an interview with the *Asahi Shimbun*, a Japanese news service in March 2013, the outgoing Governor of the Bank of England, Mervyn King, said: 'I think governments in the industrialized world do need to understand the importance of supply-side policies, such as deregulation, in order to ensure that demand will recover, not just in order to make the economy more efficient, but as a mechanism for supporting current demand.' Around the same time, the *New Statesman* magazine published a list of 20 recommendations for the Spring 2013 Budget for the UK Chancellor. Part of the recommendations included some supply-side reforms such as a Small Business Incentive Scheme which included exemptions from regulations for small businesses, such as not having to pay the minimum wage for those under 21, abolishing national pay bargaining in the public sector, simplifying planning legislation and encouraging local councils to cooperate in identifying sites for new Garden Cities. Such calls for a renewed focus on supply-side policies might suggest that the fiscal and monetary options are indeed

exhausted and that the solution to the problems of sluggish growth in many European countries and in the UK, lie in the longer-term restructuring of the economy to improve capacity.

Questions

- 1 Why have some governments had to adopt austerity policies to cut back borrowing and shrink the size of the deficit even though their economies have been either in recession or barely growing?
- 2 What is meant by 'structural deficiencies' in an economy?
- 3 In the UK in 2012–2013, employment rates were rising and inflation remained above target. Using an AS/AD diagram which utilizes the new Keynesian AS curve, explain what you think might be happening to the economy as aggregate demand rises.
- 4 Given your answer to Question 3 above, how would supply-side policies remedy the situation?
- 5 Look at the supply-side reforms suggested by the *New Statesman*. Explain how these reforms, if implemented, would help boost the supply-side of the economy?

Reform of planning laws would make it easier to redevelop derelict land and would be one example of a supply-side policy which might help to stimulate the economy.



SUMMARY

- Supply-side policies became a focus of many governments in the 1980s and into the 1990s.
- The focus is on policies designed to shift the AS curve to the right.
- Shifting the AS curve to the right can lead to sustained economic growth which increases national income but keeps prices stable countering the production bottlenecks that can occur if the focus is on increasing aggregate demand through fiscal or monetary policy.
- Supply-side policies can take two main forms – market-orientated policies or interventionist policies.
- Market-orientated policies are designed to free up markets so that they allocate resources more effectively and efficiently.
- Interventionist policies are designed to improve the working of the market and which require government intervention to overcome the public good element of the investment in infrastructure education, training and research and development.

QUESTIONS FOR REVIEW

- 1 What is meant by ‘sustainable economic growth’?
- 2 Explain the shape of the new Keynesian AS curve.
- 3 What are the two ways in which the output gap can be calculated and what is the difference between them?
- 4 What is the relevance of the size of the output gap in policy making?
- 5 Using an AS/AD diagram, show how a rise in AD could lead to a rise in national income but with no change in the price level.
- 6 What are the two main types of supply-side policy and what is the difference between them?
- 7 Why do supply-siders emphasize the importance of marginal tax rates in relation to incentives?
- 8 What is the Laffer curve?
- 9 Why do flexible labour markets improve the efficiency and capacity of the economy?
- 10 Why is it necessary for governments to invest in education, training and research and development?

PROBLEMS AND APPLICATIONS

- 1 Using the new Keynesian AS curve, explain why when the output gap is very large the AS curve has an almost horizontal slope.
- 2 Why might governments find it difficult to accurately measure the size of the output gap?
- 3 Assume that a government is successful in shifting the AS curve to the right. Show what happens to the Phillips curve in such a situation.
- 4 What sort of trade-offs do governments face if a decision was being considered to reduce the level of welfare benefits as part of a market-orientated supply-side policy?
- 5 Governments who have implemented major cuts in higher rates of income tax have been accused of giving more money to the rich at the expense of the poor. Construct an argument to counter such an accusation.
- 6 What sort of assumptions underlie the policy that cuts in income taxes would increase the supply of labour hours offered in an economy?
- 7 What are the potential costs and benefits of policies designed to improve the flexibility of labour markets?
- 8 Why do governments find it difficult to reduce spending?
- 9 A government announces a decision to increase investment in spending on higher education with the intention of increasing the participation rate in higher education by young people to 45 per cent (i.e. that 45 per cent of school leavers choose to go to university).
 - a. Does it matter what sort of degrees these young people do at university?
 - b. To what extent is the proportion of young people going on to university an important factor in improving the quality of human capital?
 - c. What else might the government need to put in place to ensure that the policy was a longer-term success (hint: will there be appropriate jobs for young people when they graduate)?
- 10 How would a government be able to tell whether a policy of increasing private sector involvement in public sector services provision actually lead to a more efficient outcome compared to having those same services provided by the public sector?

PART 15

INTERNATIONAL

MACROECONOMICS

36

COMMON CURRENCY AREAS AND EUROPEAN MONETARY UNION

During the 1990s, a number of European nations decided to give up their national currencies and use a new, common currency called the *euro* by joining European Economic and Monetary Union (EMU). Why did these countries decide to do this? What are the costs and advantages of adopting a common currency among a group of countries? Is it optimal for Europe to have a single currency? In this chapter we'll look at some of these issues, drawing on our macroeconomics tool box that we've been developing over the past several chapters.

First, a definition: a **common currency area** is a geographical area throughout which a single currency circulates as the medium of exchange. Another term for a common currency area is a *currency union*, and a closely related phenomenon is a *monetary union*: a monetary union is, strictly speaking, a group of countries that have adopted permanently and irrevocably fixed exchange rates among their various currencies. Nevertheless, the terms common currency area, currency union and monetary union are often used more or less interchangeably, and in this chapter we'll follow this practice.

common currency area a geographical area throughout which a single currency circulates as the medium of exchange

Usually we speak of common currency areas when the people of a number of economies, generally corresponding to different nation states, have taken a decision to adopt a common currency as their medium of exchange, as was the case with the European monetary union. Let's start by taking a closer look at EMU and its currency, the euro.

THE EURO

There are currently 18 countries that have joined **European Economic and Monetary Union**, or EMU. (Note that 'EMU' stands for 'Economic and Monetary Union', not European Monetary Union, as is often supposed.) The countries that currently form the euro area are Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Luxembourg, Malta, the Netherlands, Portugal, Slovakia, Slovenia and Spain (informally known as 'euroland' but more correctly as the euro area). The move towards a single European currency has a very long history. The main landmarks in its formation, started in 1992 with the Maastricht Treaty (formally known as the Treaty on European Union), which laid down (among other things) various criteria for being eligible to join the proposed currency union. In order to participate in the new currency, member states had to meet strict criteria such as a government budget deficit of less than 3 per cent of GDP, a government debt-to-GDP ratio of less than 60 per cent, combined with low inflation and interest rates close to the EU average. The Maastricht Treaty also laid down a timetable for the introduction of the new single currency and rules concerning the setting up of a European Central Bank (ECB). The ECB actually came into existence in June 1998 and forms, together with the national central banks of the countries making up the common currency area, the European System of Central Banks (ESCB), which is given responsibility for ensuring price stability and implementing the single European monetary policy.

European Economic and Monetary Union the European currency union that has adopted the euro as its common currency

The single European currency – the euro – officially came into existence on 1 January 1999 when 12 countries adopted it (although Greece did not join the EMU until 1 January 2001). On this date, exchange rates between the old national currencies of euro area countries were irrevocably locked and a few days later the financial markets began to trade the euro against other currencies such as the US dollar, as well as to trade securities denominated in euros.

The period from the beginning of 1999 until the beginning of 2002 was a transitional phase, with national currencies still circulating within the euro area countries and prices in shops displayed in both euros and local currency. On 1 January 2002 the first euro notes and coins came into circulation and, within a few months, the switch to the euro as the single medium of exchange was complete throughout the euro area.

The formation of EMU was an enormously bold step for the 12 countries initially involved. Most of the national currencies that have been replaced by the euro had been in circulation for hundreds if not thousands of years. Why did these countries deem it so important to abandon these currencies and adopt a single, common currency? Undoubtedly, part of the answer to this question lies in the realm of politics as much as in the realm of economics. However, there was also a belief that having a common European currency would help 'complete the market' for European goods, services and factors of production that had been an on-going project for much of the post-war period. More generally, the costs and benefits of adopting a common currency can be analysed within the framework of macroeconomic theory. Let's start by taking a look at the euro and its relationship to the Single European Market.

THE SINGLE EUROPEAN MARKET AND THE EURO

Following the devastation of two World Wars in the first half of the 20th century, each of which had initially centred on European conflicts, some of the major European countries (in particular France and Germany) expressed a desire to make further wars impossible between them through a process of strong economic integration that, it was hoped, would lead to greater social and political harmony. This led to the

development of the European Economic Community (EEC) – now referred to as the **European Union**, or EU. Initially the EU consisted of just six countries: Belgium, Germany, France, Italy, Luxembourg and the Netherlands. In 1973, Denmark, Ireland and the United Kingdom joined. Greece joined in 1981, Spain and Portugal in 1986, and Austria, Finland and Sweden in 1995. In 2004 the biggest ever enlargement took place with ten new countries joining. Croatia became a member state in July 2013 and at the time of writing there are five ‘candidate countries’ seeking membership. These are Turkey, the Former Yugoslav Republic of Macedonia, Iceland, Serbia and Montenegro. Albania, Bosnia and Herzegovina and Kosovo are potential candidates. The official website of the European Union defines the EU as ‘a family of democratic European countries, committed to working together for peace and prosperity’.

European Union a family of democratic European countries committed to working together for peace and prosperity

The EU has certainly been successful in its original central aim of ensuring European peace: countries such as France, England, Germany, Italy and Spain who have been at war with each other on and off for centuries now work together for mutual benefit. This has led to greater emphasis being given to the EU’s second objective – namely prosperity – and, to this end, a desire to create a **Single European Market** (SEM) throughout which labour, capital, goods and services can move freely. As member states got rid of obstacles to trade between themselves, it was argued, companies would start to enjoy economies of scale as they expanded their market across Europe, just as US companies enjoy economies of scale as they expand across American states. At the same time, inefficient firms would be exposed to more cross-border competition, either forcing them out of business or forcing them to improve their efficiency. The aim was to provide businesses with an environment of fair competition in which economies of scale could be reaped and a strong consumer base developed from which they could expand into global markets. Households, on the other hand, would benefit from lower prices, greater choice of goods and services, and work opportunities across a wide area, while the economy in general would benefit from the enhanced economic growth that would result.

Single European Market a (still not complete) EU-wide market throughout which labour, capital, goods and services can move freely

Early steps towards the creation of the SEM included the abolition of internal EU tariff and quota barriers in 1968 and a movement towards greater harmonization in areas such as indirect taxation, industrial regulation, and in common EU-wide policies towards agriculture and fisheries.

Nevertheless, it proved difficult to make progress on the more intangible barriers to free movement of goods, services, capital and labour. For example, even though internal tariffs and quotas had been abolished in the EU, local tax systems and technical regulations on goods and services still differed from country to country so that it was in practice often difficult to export from one country to another. Thus, a car produced in the UK might have to satisfy a certain set of emission and safety requirements in one European country and another set of requirements in another EU country. Or a qualified engineer might find that her qualifications, obtained in Italy, were not recognized in Germany. The result was that during the 1970s and early 1980s, growth in the EU member states began to lag seriously behind that of international competitors – especially the United States and Japan. Therefore, in 1985 a discussion document (in the jargon, a ‘White Paper’) was produced by the European Commission that subsequently led to a European Act of Parliament – the 1986 Single European Act. This identified some 300 measures that would have to be addressed in order to complete the Single European Market and set 31st December 1992 as the deadline for completion. The creation of the SEM was to be brought

about by EU Directives telling the governments of member states what changes needed to be put into effect in order to achieve four goals:

- The free movement of goods, services, labour and capital between EU member states.
- The approximation of relevant laws, regulations and administrative provisions between member states.
- A common, EU-wide competition policy, administered by the European Commission.
- A system of common external tariffs implemented against countries who are not members of the EU.

Over 20 years on from the Single European Act, the SEM is still far from complete. In particular, there still exist between EU members strong differences in national fiscal systems, while academic and professional qualifications are not easily transferable and labour mobility across EU countries is generally low. Some of the reasons for this are hard to overcome: language barriers and relative levels of economic development hamper the movement of factors and member states continue to compete with one another economically, at times seeking their own national interest rather than the greater good of the EU.

Nevertheless, the years between 1985 and 1992 did see some important steps in the development of the SEM and the resulting achievements of the SEM project were not negligible: the European Commission estimates that the SEM helped create 2.5 million new jobs and generated €800 billion in additional wealth in the ten years or so following 1993.

In the context of the Single European Market project, therefore, the creation of a single European currency was seen as a final step towards 'completing the market', by which was meant two things: (a) getting rid of the transaction costs from intra-EU trade that result from different national currencies (and which act much as a tariff) and (b) removing the uncertainty and swings in national competitiveness among members that result from exchange rate movements. Before EMU, most EU countries participated in the Exchange Rate Mechanism (ERM), which was a system designed to limit the variability of exchange rates between members' currencies. However, the ERM turned out not to be a viable way of reducing volatility in the exchange rate and, in any case, had no effect on the transaction costs arising from bank charges associated with changing currencies when engaging in intra-EU trade.

It is clearly important, therefore, to see EMU within a broader European framework and, in particular, the Single European Market project. Nevertheless, the benefits of adopting a single currency across a geographical area can be analysed more generally using macroeconomic theory. Moreover, these benefits must be weighed against the costs of joining a common currency area.

THE BENEFITS AND COSTS OF A COMMON CURRENCY

Benefits of a Single Currency

Elimination of Transaction Costs One obvious and direct benefit of a common currency is that it makes trade easier between members and, in particular, there is a reduction in the transaction costs involved in trade between members of the common currency area. When a German company imports French wine, it no longer has to pay a charge to a bank for converting German marks into French francs with which to pay the wine producer, it can just pay in euros. Of course, the banking sector loses out on the commission it used to charge for converting currencies, but this does not affect the fact that the reduction in transaction costs is a net gain. This is because paying a cost to convert currencies is in fact a deadweight loss in the sense that companies pay the transaction cost but get nothing tangible in return. OK, so the banks were getting commission before and this was used to employ people who worked on currency transactions, but these people can now be used more productively in the economy, making everyone better off.

Reduction in Price Discrimination It is sometimes argued that a second albeit indirect gain to the members of a common currency area results from the reduction in price discrimination that should ensue when there is a single currency. If goods are priced in a single currency it should be much harder to disguise price differences across countries. As we have discussed, price discrimination involves a deadweight loss to society, so this is a further gain from a single currency. This argument assumes that the transparency in

prices that results from a common currency will lead to arbitrage in goods across the common currency area: people will buy goods where they are cheaper (tending to raise their price in that location) and reduce their demand for goods where they are more expensive (tending to reduce the price in that location).

Overall, however, EMU seems unlikely to bring an end to price discrimination across euro area countries. For items like groceries, having a single currency is unlikely to be much of an impetus to price convergence across the common currency area because of the large transaction costs (mainly related to travelling) involved in arbitraging, relative to the prices of the goods themselves. Would you travel to another country to do your grocery shopping if you knew the prices there were lower? Perhaps you might – when it is easy to travel this does happen. Doing so across the whole of the EU means that the effects are necessarily more limited. On the other hand, big ticket items like household appliances and electronic goods, where the transaction costs may be lower as a percentage of the price of the good in question, are also unlikely to be arbitraged heavily across national borders by consumers because of their durable nature and the need for confidence in after-sales service. Would you want to buy a fridge-freezer in a foreign country and run the risk of having to take it back or get someone from there to fix it for you if it goes wrong? In addition, the fact that different countries in the EU still have different plugs and power systems, which are difficult to harmonize, might also reduce the overall impact.

Reduction in Foreign Exchange Rate Variability A third argument relates to the reduction in exchange rate variability and the consequent reduction in uncertainty that results from having a single currency. Exchange rates can fluctuate substantially on a day-to-day basis. Before EMU, when a German supermarket imported wine from France to be delivered, say, three months later, it had to worry about how much a French franc would be worth in terms of German marks in three months' time and therefore what the total cost of the wine would be in marks. This uncertainty might deter the supermarket company from importing wine at all, and instead lead it to concentrate on selling German wines, thereby foregoing the gains from trade and reducing economic welfare. The supermarket could have eliminated the uncertainty by getting a bank to agree to sell the francs at an agreed rate against marks to be delivered three months later (an example of a forward foreign exchange contract). But the bank would charge for this service, and this charge would be equivalent to a tariff on the imported wine and so this again would represent a deadweight loss to society.

The reduction in uncertainty arising from the removal of exchange rate fluctuations may also affect investment in the economy. This would clearly be the case for companies that export a large amount of their output to other euro area countries, since less uncertainty concerning the receipts from its exports means that it is able to plan for the future with less risk, so that investment projects such as building new factories appear less risky. An increase in investment will benefit the whole economy because it is likely to lead to higher economic growth.

Costs of a Single Currency

The major cost to an economy in joining a common currency area relates to the fact that it gives up its national currency and thereby gives up its freedom to set its own monetary policy and the possibility of macroeconomic adjustment coming about through movements in the external value of its currency. Clearly, if the nations of the euro area have only one money, they can have only one monetary policy, which is set and implemented by the ECB. This must be the case because, since there is only one currency, it's not possible to have a different set of interest rates in different countries. Why is this a potential problem?

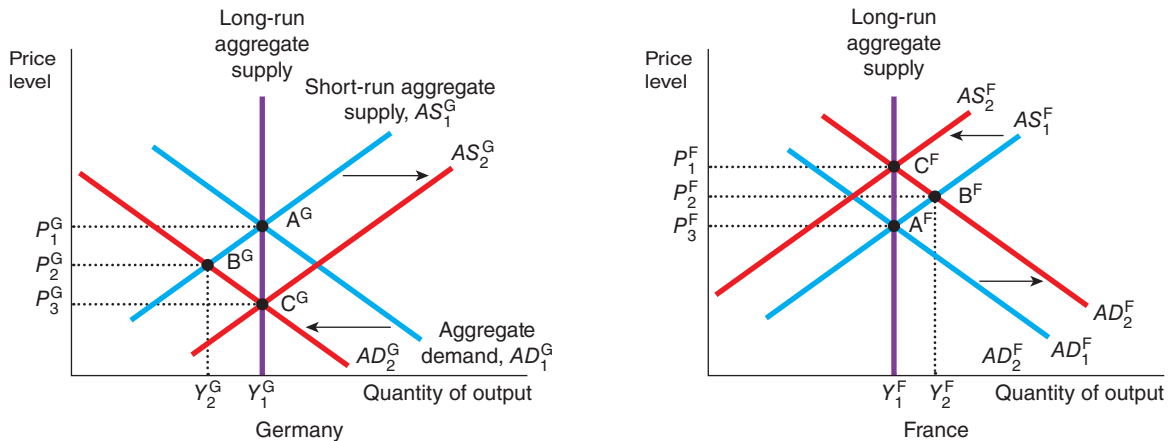
Suppose, for example, that there is a shift in consumer preferences across the common currency area away from goods and services produced in one country (Germany, say) and towards goods and services produced in another country (France, say). This situation is depicted in Figure 36.1, which shows a leftwards shift in the German short-run aggregate demand curve and a rightwards shift in the French short-run aggregate demand curve. What should policymakers in France and Germany do about this? One answer to this is, nothing: in the long run, each economy will return to its natural rate of output. In Germany, this will occur as the price level falls and wages, prices and perceptions adjust. In particular, as unemployment rises in Germany, wages eventually begin to fall. Lower wages reduce firms' costs and so, for any given price level, the amount supplied will be higher. In other words, the German short-run aggregate supply curve will

shift to the right, until eventually it intersects with the new short-run aggregate demand curve at the natural rate of output. The opposite happens in France, with the short-run aggregate supply curve shifting to the left. The adjustment to the new equilibrium levels of output is also shown in Figure 36.1.

FIGURE 36.1

A Shift in Consumer Preferences Away from German Goods Towards French Goods

The German fall in aggregate demand leads to a fall in output from Y_1^G to Y_2^G , and a fall in the price level from P_1^G to P_2^G . The increase in French aggregate demand raises output from Y_1^F to Y_2^F . Over time, however, wages and prices will adjust, so that German and French output return to their natural levels, Y_1^G and Y_1^F , with lower prices in Germany, at P_3^G , and higher prices in France, at P_3^F .



Note that, if Germany and France had maintained their own currencies and a flexible foreign exchange rate, then the short-term fluctuations in aggregate demand would be alleviated by a movement in the exchange rate: as the demand for French goods rises and for German goods falls, this would increase the demand for French francs and depress the demand for German marks, making the value of francs rise in terms of marks in the foreign currency exchange market. This would make French goods more expensive to German residents since they now have to pay more marks for a given number of French francs. Similarly, German goods become less expensive to French residents. Therefore, French net exports would fall, leading to a fall in aggregate demand. This is shown in Figure 36.2, where the French aggregate demand schedule shifts back to the left until equilibrium is again established at the natural rate of output. Conversely – and also shown in Figure 36.2 – German net exports rise and the German aggregate demand schedule shifts to the right until equilibrium is again achieved in Germany.

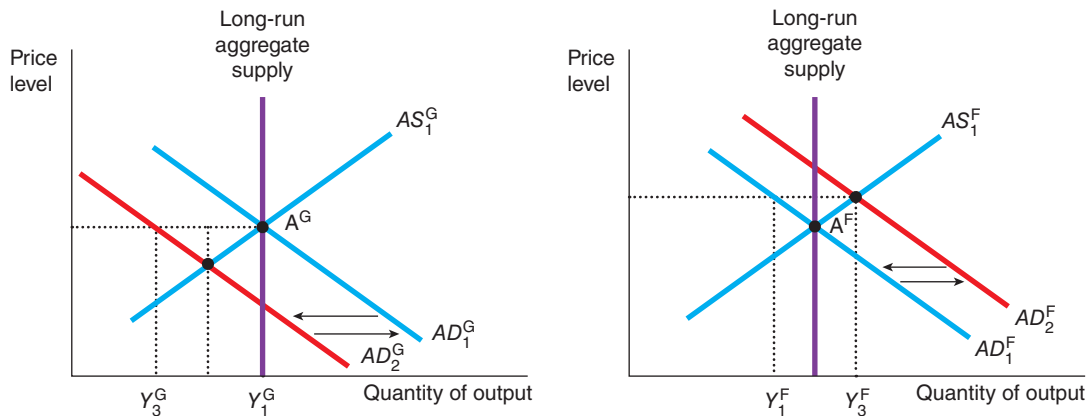
In a currency union, however, this automatic adjustment mechanism is not available, since, of course, France and Germany have the same currency (the euro). The best that can be done is to wait for wages and prices to adjust in France and Germany so that the aggregate supply shifts in each country, as in Figure 36.1. The resulting fluctuations in output and unemployment in each country will tend to create tensions within the monetary union, as unemployment rises in Germany and inflation rises in France. German policymakers, dismayed at the rise in unemployment, will favour a cut in interest rates in order to boost aggregate demand in their country, while their French counterparts, worried about rising inflation, will be calling for an increase in interest rates in order to curtail French aggregate demand. The ECB will not be able to keep both countries happy. Most likely, it will set interest rates higher than the German desired level and lower than the French desired level. The ECB pursues an inflation targeting strategy, and the inflation rate it targets is based upon a consumer prices index constructed as an average across the euro area. If a country's inflation rate (or expected inflation rate) is below the euro area average, the ECB's monetary policy will be too tight for that country; if it is above the average, the ECB's monetary policy will

be too loose for it. All that is possible is a 'one size fits all' monetary policy. It is for this reason that entry to the eurozone is restricted to those countries that can meet the criteria outlined above where inflation and interest rates are close to the EU average.

FIGURE 36.2

A Shift in Consumer Preferences with Flexible Exchange Rates

The fall in German aggregate demand leads, before prices have had time to adjust, to a fall in output from Y_1^G to Y_3^G . However, because this is due to a fall in net foreign demand, the value of the German currency falls, making German goods cheaper abroad. This raises net exports and restores aggregate demand. The converse happens in France: the increase in net foreign demand raises the external value of the French currency, making French goods more expensive abroad and choking off aggregate demand to its former level.



SELF TEST List and discuss the key costs and benefits of joining a currency union.

THE THEORY OF OPTIMUM CURRENCY AREAS

Optimum currency area (OCA) theory attempts to set down a set of criteria for a group of countries such that, if the criteria were satisfied, then it would in some sense be 'optimal' for the countries to adopt a common currency. The qualifier 'optimal' here refers to the ability of each of the countries to limit the costs of monetary union and enhance the benefits. It is generally used loosely, since there is no way for certain of ensuring whether it is indeed optimal for a group of countries to form a currency union and, more often than not, countries will fulfil some but not all of the OCA criteria.

optimum currency area a group of countries for which it is optimal to adopt a common currency and form a currency union

Characteristics that Reduce the Costs of a Single Currency

Consider first the characteristics of a group of countries that would reduce the costs of adopting a common currency. As we have discussed, the main cost to participating in a monetary union is the loss of monetary policy autonomy for the individual countries concerned, as well as ruling out the possibility

of macroeconomic adjustment through exchange rate movements. One way in which the economic (and political) tensions arising from the loss of the exchange rate instrument and the imposition of a 'one-size-fits-all' monetary policy will be alleviated is if the economies in question move rapidly to long-run equilibrium following a macroeconomic shock: since we know there is only a short-run trade-off between inflation and unemployment, the faster the economies concerned can get to the long run – in other words, return to their natural rates of output and unemployment – the better. This speed of adjustment to long-run equilibrium will be high if there is a high degree of wage flexibility in the common currency area, and/or if there is a high degree of labour mobility.

Another way in which tensions across the common currency area would be alleviated would be if all countries in the currency union were prone to the same kind of demand shocks (e.g. if aggregate demand fell in all countries simultaneously), since then each would favour similar macroeconomic policy decisions (e.g. a reduction in interest rates).

We consider each of these types of characteristics in turn.

Real Wage Flexibility Suppose there is a high degree of wage flexibility in each of the member countries, so that wages respond strongly to rises and falls in unemployment. This means that the adjustment to long-run equilibrium, as shown in Figure 36.1, occurs very quickly. In our example, the shift in aggregate demand in Germany leads to falling wages, so that firms make more profit for any given level of prices and the aggregate supply curve shifts to the right and Germany returns to the natural rate of output. If wages are very flexible, this adjustment may be very rapid, so that the short run is very short indeed. Similar for France: the rightwards shift in aggregate demand leads to rapidly rising wages and firms find it less profitable to produce any given level of output, so that the supply curve shifts leftwards and a new long-run equilibrium is established at the natural rate of output. Hence, by compressing the short run, tensions across the monetary union are ironed out very quickly.

Note that it is the real wage that is of importance here: it is real wages that must adjust in order to affect the aggregate supply curve by making it more (or less) profitable for firms to produce a given level of output at any given level of prices.

Labour Mobility Alternatively, suppose that labour is highly mobile between the member countries of the currency union: unemployed workers in Germany simply migrate to France and find a job. Again, the macroeconomic imbalance is alleviated, since unemployment in Germany will fall as many of the unemployed have left the country, and inflationary wage pressures in France decline as the labour force expands with the migrants from France. Therefore, it is clear that labour mobility may in some measure cushion a currency union from **asymmetric shocks**, a situation where changes in aggregate demand and/or supply differ from one country to another.

asymmetric shocks a situation where changes in aggregate demand and/or supply differ from one country to another

Capital Mobility Sometimes economists argue that capital mobility can also compensate for the loss of monetary autonomy and the absence of exchange rate adjustment among the members of a common currency area. A distinction should be made here between physical capital (plant and machinery) and financial capital (bonds, company shares and bank loans). In terms of cushioning a currency union from asymmetric shocks, movements in physical capital can help by expanding productive capacity in countries experiencing a boom, as firms in other member countries build factories there. However, given the long lags involved in the installation of plant and equipment, physical capital mobility is likely to be helpful mainly for narrowing persistent regional disparities rather than offsetting short-term shocks.

The mobility of *financial* capital may be more useful in cushioning economies from short-term output shocks. For example, residents of a country experiencing a recession may wish to borrow money from the residents of a country experiencing a boom in order to overcome their short-term difficulties. In our two-country example, German residents would effectively borrow money from French residents in order to make up for their temporary fall in income. Clearly, this would require that German residents can easily

borrow from French residents through the capital markets, so that financial capital mobility will be highest between countries whose capital markets are highly integrated with one another. For example, if a bank has branches in more than one country of a currency union, then borrowing and lending between growth and recession countries will be more or less automatic, as residents in the growth country increase the money they are holding in the bank as their income goes up and residents of the country in recession increase their overdrafts (or reduce their money holdings) as their income goes down.

We can relate this discussion back to the notion of permanent income, and to the market for loanable funds. Recall that a family's ability to buy goods and services depends largely on its permanent income, which is its normal, or average income, since people tend to borrow and lend to smooth out transitory variations in income. Now, when an aggregate demand shock adversely affects the German economy, a large amount of German households will see their transitory income fall and will want to borrow in order to increase income back up to the permanent or normal level. But since many German households are now doing this at the same time, if borrowing is restricted to German financial markets, this will tend to raise interest rates and generally make borrowing more difficult. If the market for loanable funds is restricted to the domestic market, then we might expect the supply of loanable funds to decrease in a recession and the demand to increase, raising interest rates. The resulting rise in interest rates may even make the recession worse by reducing investment.

On the other hand, in France, the economic boom means that many households are experiencing income levels above their permanent or average level, and so will tend to increase their saving. Now, if the German households can borrow from the French households – if the market for loanable funds covers both France and Germany – they can both consume at a level consistent with their normal or average levels of income with less of an effect on interest rates. There will be an increase in the supply of loanable funds because French residents are saving more and this will partly or even wholly offset the increase in the demand for loanable funds arising from German residents who want to borrow more. When the German economy comes out of recession and goes into recovery, German households can then repay the loans.

Of course, although we have discussed only bank loans, there are other forms of financial capital, such as bonds and company shares, but the principle of the recessionary economy being able to obtain funds from the booming economy remains the same. In effect, therefore, financial capital market integration across countries allows households to insure one another against asymmetric shocks so that the variability of consumption over the economic cycle can be reduced.

Symmetric Macroeconomic Shocks Note that, in describing the costs of belonging to a monetary union, we have used the example of a positive demand shock in one country and a simultaneous negative demand shock in another. A similar analysis would have followed if we had simply allowed either a positive or a negative demand shock in one country and no shock at all in the other country. The central point was that the demand shock was asymmetric in the sense that it impacted differently on different members of the currency union, requiring different short-run policy responses. Clearly, if the shock were symmetric there would be no problem. If, for example, aggregate demand rose simultaneously in all member countries, increasing expected future inflation, then a policy of raising interest rates would be welcomed by all members of the monetary union. This would be the case if the economic cycles of each of the countries making up the currency were synchronized in the sense that the various economies tended to enter recession at the same time and enter the recovery phase of the cycle at the same time, so that disagreements about the best interest rate policy are less likely to occur.

Characteristics that Increase the Benefits of a Single Currency

High Degree of Trade Integration The greater the amount of trade that takes place between a group of countries – i.e. the greater the degree of trade integration – the more they will benefit from adopting a common currency. One of the principal benefits of a currency union, and the most direct benefit, is the reduction in transaction costs that are incurred in trade transactions between the various countries when there is a constant need to switch one national currency into another on the foreign currency exchange market. Clearly, therefore, the greater the amount of international trade that is carried out between member countries – and therefore the greater the amount of foreign currency transactions – the greater the reduction in transaction costs that having a common currency entails.

The reduction in exchange rate volatility – another benefit of a currency union – will also clearly be greater, the greater is the degree of intra-union trade, since more firms will benefit from knowing with certainty exactly the revenue generated from their sales to other currency union members, rather than having to bear the uncertainty associated with exchange rate fluctuations.

SELF TEST What is meant by an optimum currency area? • List and discuss the key characteristics of an optimum currency area.

IS EUROPE AN OPTIMUM CURRENCY AREA?

Having determined what characteristics of a group of countries would make the benefits of a single currency stronger and the costs weaker, we can take a closer look to see whether Europe – and in particular the group of 18 countries that comprise the euro area – forms an optimum currency area.

Trade Integration

The degree of trade integration can be assessed by looking at imports from and exports to other EU countries expressed as a percentage of GDP. The degree of trade integration across Europe is quite variable, but nevertheless on average quite high – with the notable exception of Greece. The degree of European trade integration has been rising over time, in nearly every country in the EU but the integration has been more marked in some countries such as Austria than others such as the UK and Italy.

This has led some economists to argue that some of the criteria for an optimum currency area – such as a high degree of trade integration – may actually be endogenous: actually being a member of a currency union may enhance the degree of trade done between members of the union, precisely because of the decline in transaction costs in carrying out such trade.

Overall many – if not all – European countries have gained a great deal from the reduction in transaction costs in international trade as a result of the single currency. Indeed, these gains have been estimated at about one quarter to one half of one per cent of euro area GDP. This may not sound massive, but remember that transaction costs are a deadweight loss. Moreover, the gains are not one-off: they accrue continuously as long as the single currency persists, since they would have to be paid in the absence of the currency union. They therefore become cumulative. In addition, if the degree of euro area trade integration tends to rise over time as a result of the single currency, as some economists have suggested, then the implicit gain from not having to pay transaction costs also rises over time.

The other, indirect benefit of a single currency when there is a high degree of trade integration, follows from the reduction in uncertainty associated with doing away with the volatility in the exchange rates between members' national currencies (since those currencies are replaced with a common currency). These gains are hard to quantify, but it is not incorrect to suggest that they are not negligible for the euro area.

Real Wage Flexibility

A great deal of research has been done on real wage flexibility in Europe and virtually all of it concludes that continental European labour markets are among the most rigid in the world, while the UK labour market, at least since the 1980s, has become one of the most flexible. One reason for this is the fact that all European Union countries have minimum wage laws, although this is not the whole story since the UK also has a minimum wage. Perhaps a more important reason is the high degree of collective wage bargaining that is common in continental Europe – i.e. wage agreements that cover a large number of workers. Figures on the degree of unionization of the labour force are quite deceptive in this sense: for example, in the early 2000s, in the UK, Italy and Germany about 30 per cent of the workforce belonged to a trade union, while in France the figure was around 10 per cent. Continental European unions, however,

often have collective bargaining and other workplace rights that UK trade unionists can only envy. The power of unions and the inflexibility of the labour market is something that has been a concern in France for some time and in 2013, an agreement was reached between the government and unions on reforms to the country's complex and strict labour laws. The agreement was greeted with mixed responses with some saying it did not go far enough and others suggesting the agreement marked the end of workers' rights in France.

The introduction of the single European currency may also have had a negative effect on European wage flexibility, since many European collective wage agreements between workers and a firm in one country will also often extend to the firm's workforce in other European countries, and a single currency brings transparency in wage differences across countries, as well as price transparency. To return again to our example of a negative demand shock in Germany and a positive shock in France, a company with employees in both countries would find it hard to reduce real wages in Germany while raising them in France.

Furthermore, European labour law is generally very much more restrictive in many continental European countries than it is in the UK or the USA, as is the level of payroll taxes, so that a firm's costs of either reducing the workforce or increasing it can be very high. This means that, even if there were movements in the real wage, firms would be slow to expand or contract their output in response, so that shifts in aggregate supply will be slow to come about.

On the whole, therefore, adjustment to asymmetric shocks through real wage movements is unlikely to be significant in the euro area.

Labour Mobility

Labour is notoriously immobile across European countries, at least if one rules out migration from the newer eastern European members of the EU such as Poland, Romania and Bulgaria, and considers just the 18 euro area countries or these plus Denmark, Sweden and the UK. In part this may perhaps be attributed to differences in language, culture and other social institutions across Europe that make it difficult for workers to migrate. However, it seems that European workers are also very loath to move location even within their own countries. Indeed, the degree of labour mobility as measured by the percentage of the workforce that moves geographical location over any given period, is much lower within any particular European country than it is within the United States, and is even lower between the euro area countries. Europe therefore scores very low on this optimum currency area criterion.

Financial Capital Mobility

In discussing financial capital mobility, a distinction must be made between the wholesale and the retail capital markets. The wholesale financial markets are the capital markets in which only financial institutions such as banks and investment funds operate, as well as very large corporations, while the retail financial markets (such as high street banks) are those open to individual households and to small and medium-sized corporations. Prior to the introduction of the euro, financial integration among euro area countries was probably quite low, in both the wholesale and retail sectors. However, following the introduction of the euro, integration of the wholesale financial markets has increased dramatically. In particular, a liquid euro money market with single interbank market interest rates was established so that a bank in, say, Luxembourg can now borrow euros just as easily and at the same rate of interest from another bank in Frankfurt as it can from a bank located in the same street in Luxembourg. In the government bond market, the degree of market integration is also high, and this is shown by the fact that the interest rates on government bonds of the different euro area countries are very close to one another and tend to move very closely together. On the other hand, the integration of retail market products, such as loans to households and small and medium-sized enterprises, is lagging behind compared with the wholesale market products. This becomes evident from persistent cross-country differences in bank lending rates and the rather limited cross-border retail banking activity. Indeed, national banking sectors have remained largely segregated with only marginal cross-border penetration.

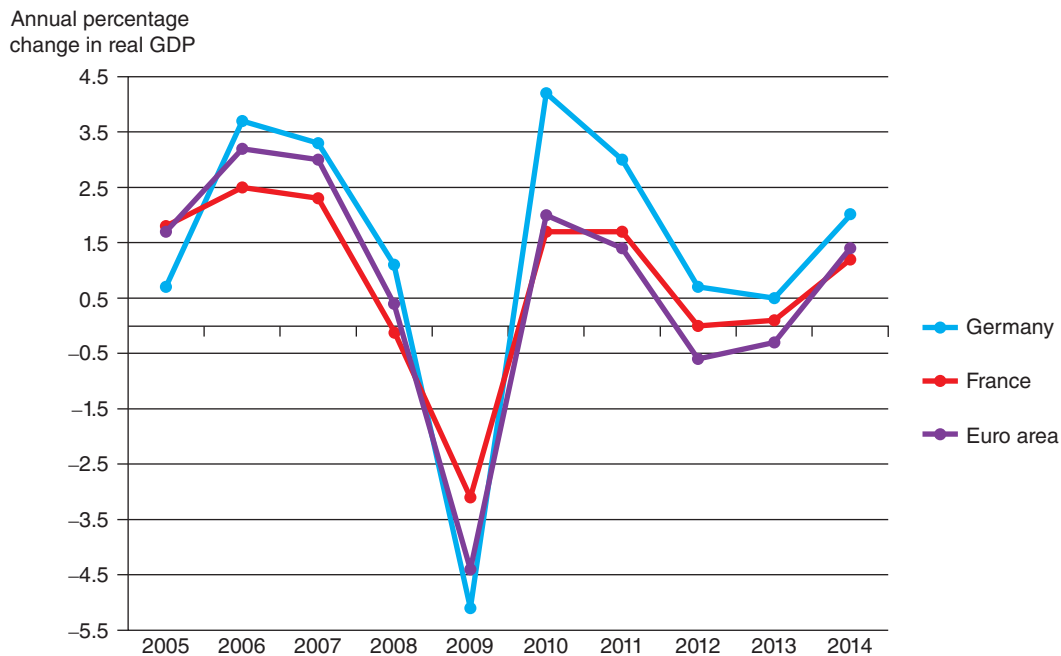
Symmetric Demand Shocks

The economic cycle across the countries of the euro area does appear to be positively correlated, in the sense that the timing of strong growth and downturns appear to be very close. In Figure 36.3 we have graphed data on annual growth rates in real GDP for France, Germany and the euro area as a whole for every year from 2005 to 2014. Clearly, the movements in growth rates over this period for France and Germany and for the whole euro area are very close. There is no clear example of asymmetric demand shocks impacting upon these countries as in the example we have used to illustrate the potential problems of a monetary union.

FIGURE 36.3

Growth Rates in Germany, France and the Whole Euro Area

The figure shows annual growth rates in real GDP for France, Germany and the euro area as a whole for every year from 2005 to 2014. French, German and euro area (17) growth rates (Latvia only joined the euro area in January 2014) tend to move closely together and there is no clear example of asymmetric demand shocks.

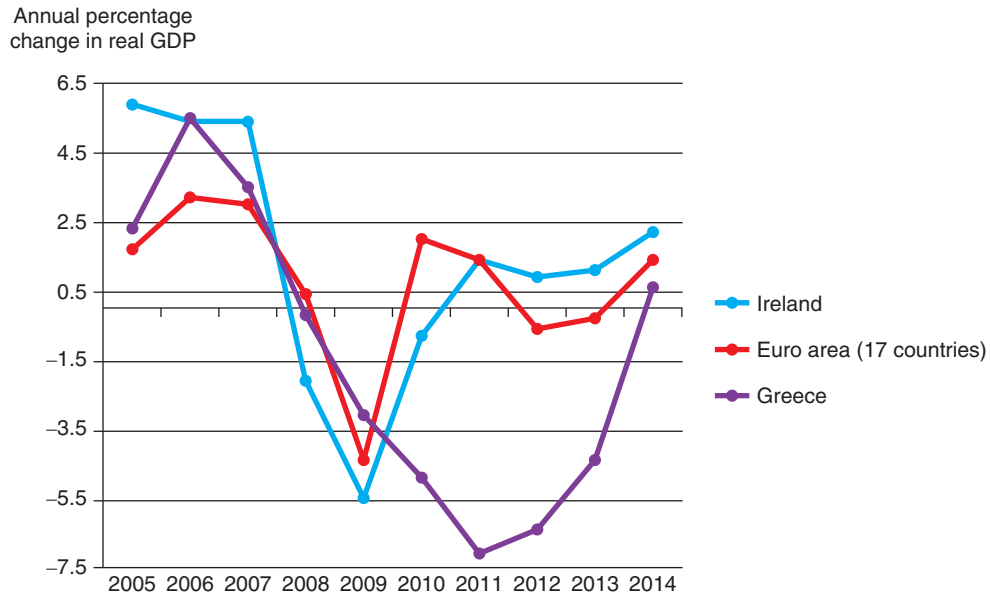


In Figure 36.4, however, we have plotted real GDP growth rates for the euro area, Greece and Ireland over the same period. The problem that is evident from this graph is not so much that the turning points do not coincide, but that Ireland and Greece's growth rate outstripped the performance of the euro area as a whole between 2005 and 2007 but then both saw considerable problems as a result of the financial crisis. Whilst Ireland has seen a recovery and is performing at a rate higher than the euro area as a whole, Greece has seen its growth shrinking throughout the period 2008–2013. Whilst the euro area as a whole saw some weak recovery in that period, Greece did not share in that recovery. Monetary policy for the euro area as a whole has not been matched to the needs of the Greek and Irish economies over the period noted.

Overall, therefore, the evidence is a little mixed, although on the whole it suggests that the problem of asymmetric demand shocks is not a great one for the current member countries of EMU. The fact that there is not strong evidence of asymmetric demand shocks at the aggregate level, however, does not rule out the possibility that there may be asymmetric shocks at other levels in the economy.

FIGURE 36.4**Growth Rates in Ireland, Greece and the Whole Euro Area**

The figure shows annual growth rates in real GDP for Ireland, Greece and for the euro area as a whole for every year from 2005 to 2014 (note, 2014 is an estimate). The growth rates do not move closely together; Ireland and Greece's growth rate was higher than that of the euro area as a whole up to 2007, but Ireland slipped into recession first and the depth of Greece's recession has been much worse than that of the euro area.



In fact, researchers have found that many of the shocks that impact upon European countries asymmetrically tend to be specific to a region or to an industry rather than to a country as a whole. This is not a problem made worse by joining a monetary union, however, since a country that experienced, say, a negative shock to one of its industries or regions would not in any case be able to deal with this using monetary or exchange rate policy without generating imbalances in its other regions or industries. The idea of a two-speed EU with the countries that have experienced debt crisis problems since the financial crisis, Portugal, Ireland, Italy, Greece and Spain (the so called 'PIIGS' countries) has highlighted that there are some countries that seem better able to weather economic storms than others. Germany, France, the Netherlands, Denmark and Sweden have all recovered from the financial crisis more quickly than have the PIIGS, albeit that the strength of the recovery has been muted. Economists have noted that the degree of integration of the euro area and the EU as a whole does impact on the ability of countries to recover from economic shocks. If EU countries rely on each other for export-led growth, then if a group of these countries face considerable fiscal difficulties and have to implement austerity measures, then the level of demand in these economies falls and this affects demand in the other economies, thus dragging down growth in all the euro area countries and the EU as a whole.

Summary: So is Europe an Optimum Currency Area?

As in many policy debates in economics, there is no clear-cut answer to this question. Certainly, many European countries have a high degree of intra-union trade and have economic cycles that are more or less synchronized. However, labour mobility and wage flexibility (and labour market flexibility in general) are low in Europe, and while the euro has increased financial market integration in the euro area's wholesale financial markets, the retail markets remain highly segregated at the national level.

Overall, therefore, if very strong differences in the economic cycle were to emerge across the euro area, the lack of independent monetary and exchange rate policy would be acutely felt. This could be a case argued in relation to the situations in Ireland and Greece as highlighted in Figure 36.4. For that reason, many economists argue that Europe – meaning the current euro area – is not an optimum currency area. Nevertheless, as we have noted in our discussion, it is possible that some of the optimum currency area criteria may be endogenous. In particular, the single currency is likely to generate even greater trade among EMU members. Given this, it is likely that the economic cycles of member currencies will become even more closely synchronized as aggregate demand shifts in one country have stronger and stronger spillover effects in other euro area countries; this may, however, take some time to synchronize and the problems the EU has faced since the financial crisis have exposed the differences in the relative strength of the economies of the countries, in the euro area in particular, which has pushed EMU to its limits.

Over time, the single currency, if it survives, may also raise labour mobility across Europe in the long run, since being paid in the same currency as in one's home country is one less issue to come to terms with when moving location to find a job. Also, with time, one would expect financial market integration to spread to the retail capital markets (indeed there is already strong pressure being exerted on the European banking industry by the European Central Bank and the European Commission for the introduction of new payment schemes for electronic credit transfers and direct debits between euro area banks, as well as for a unified framework for debit and credit cards).

Perhaps the only true test of whether the euro area is an optimum currency area (or can become one) is to see whether EMU survives in the long run. The challenges faced by the euro area since the financial crisis have put the long-term survival of the whole euro project in doubt. Significant changes to the structure of the euro, in particular financial regulation and fiscal reform, are the subject of on-going discussion and fiscal policy is the subject of our next section.

FISCAL POLICY AND COMMON CURRENCY AREAS

Our discussion so far has tended to centre on the loss of autonomy in monetary policy that is entailed in adopting a single currency among a group of countries. However, it is obvious that there is nothing in the adoption of a common currency that implies that members of the currency union should not still retain independence in fiscal policy. For instance, in our example of an asymmetric demand shock that expands demand in France and contracts aggregate demand in Germany, the French government could reduce government spending in order to offset the demand shock, while the German government could expand government spending. In fact, even if France and Germany did not make up an optimal currency because wages were sticky and labour mobility was low between the countries, national fiscal policy could, in principle, still be used to ameliorate the loss of monetary policy autonomy. Let's discuss further the issue of fiscal policy and currency unions.

Fiscal Federalism

Suppose that a currency union had a common fiscal policy in the sense of having a single, common fiscal budget covering tax and spending decisions across the common currency area. This means that fiscal policy in the currency union would work much as fiscal policy in a single national economy works, with a surplus of government tax revenue over government spending in one region used to pay for a budget deficit in another region. Return again to our example of an asymmetric demand shock that expands aggregate demand in France and contracts aggregate demand in Germany, as in Figure 36.1. Remember that there are automatic stabilizers built into the fiscal policy of an economy that automatically stimulate aggregate demand when the economy goes into recession without policymakers having to take any deliberate action. In particular, since almost all taxes are closely related to the level of economic activity in the economy, tax revenue will automatically decline in Germany as a result of the aggregate demand shock that shifts it into recession. At the same time, transfer payments in the form of unemployment benefit and other social security benefits will also rise in Germany. The opposite will be true in France,

where the automatic stabilizers will be operating in reverse as transfer payments fall and tax receipts rise with the level of economic activity. These changes will tend to expand aggregate demand in Germany and contract it in France, to some extent offsetting the asymmetric demand shock.

Now, if the governments of France and Germany have a common budget, then the increased net government revenue in France can be used to offset the reduction in net government revenue in Germany. If the resulting movements in aggregate are not enough to offset the demand shock, then the French and German governments may even go further and decide to increase government expenditure further in Germany and pay for it by reducing spending and perhaps raising taxes in France.

This kind of arrangement – a fiscal system for a group of countries involving a common fiscal budget and a system of taxes and fiscal transfers across countries – is known as **fiscal federalism**. The problem with it is that the taxpayers of one country (here France) may not be happy in paying for government spending and transfer payments in another country (in this example, Germany).

fiscal federalism a fiscal system for a group of countries involving a common fiscal budget and a system of taxes and fiscal transfers across countries

National Fiscal Policies in a Currency Union: The Free Rider Problem

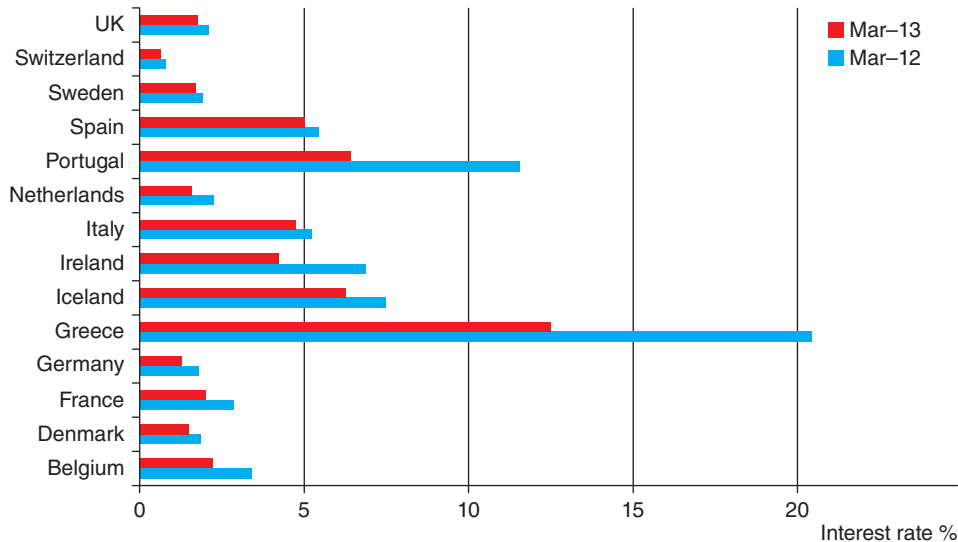
Assuming that, for political reasons, fiscal federalism is not an option open to the currency union, we still need to explore the possibility of individual members of the union using fiscal policy in order to offset asymmetric macroeconomic shocks that cannot be dealt with by the common monetary policy. In particular, in our example, what is wrong with Germany running a big government budget deficit in order to counteract the fall in aggregate demand and borrowing heavily in order to finance the deficit? One answer may lie in the effect on other members of the currency union of a rise in the debt of a member country.

Whenever a government raises its levels of debt to very high levels, there is always the possibility that the government may default on the debt. In general, this can be done in one of two ways. Where a country is not a member of a currency union and controls its own monetary policy, it can engineer a surprise inflation by a sudden increase in the money supply, so that the real value of the debt shrinks. In addition, when there is a sharp rise in the price level, this will usually be accompanied by a sharp fall in the foreign currency value of the domestic currency. This means that, valued in foreign currency, the stock of government debt will now be worth far less. Thus, the government has in effect defaulted on a large portion of its debt by reducing its value both internally and externally.

If this is not possible – for example because, as in a currency union, the country no longer enjoys monetary policy autonomy and is not able to devalue the external value of its currency (since it uses the common currency) – then the only other way of reneging on the debt is through an outright default (e.g. stopping interest payments or failing to honour capital repayments when they fall due). Generally, the financial markets are good at disciplining governments that run up large debts, by charging them high rates of interest on the debt that the government issues – after all, if you thought there was even a slight possibility that you might not get your money back if you lent it, you would want to be paid a higher rate of interest in order to compensate for that risk. Figure 36.5 shows the interest rate on ten-year government debt in a selection of EU countries in March 2013 compared to March 2012. The overall cost to governments of borrowing was lower in March 2013 reflecting the slightly more calm economic conditions that prevailed at the time compared to a year previously. The differences in the cost of borrowing is stark, however. The PIIG countries all face much higher interest rates on borrowing compared to the stronger economies of Germany, France, the Netherlands, Sweden, Denmark, the UK and Switzerland. The interest rate on Greek debt is extreme reflecting the parlous state of that country's economy. It is rarely impossible for a country to borrow money on international markets but if interest rates are higher than about 7 per cent then the cost of servicing the debt becomes unsustainable in the longer term, particularly if economic growth is weak as it has been in the PIIG countries. If the cost of borrowing is consistently above 7 per cent then it is likely that governments will be in even more trouble and become more likely to default.

FIGURE 36.5**Interest Rate on Ten-year Government Debt in Selected European Countries**

The figure shows the interest rate on ten-year government debt for a selection of European countries in March 2012 and March 2013. The cost of borrowing is generally lower for all countries in March 2013 but there is a considerable variation in the cost of borrowing. The cost tends to be much lower for those countries with stronger economies and thus less likely to default.



In the case of a monetary union, therefore, this means that excessive debt issuance by one member country will tend to force up interest rates throughout the whole of the common currency area. Although the ECB controls very short-term interest rates in the euro area through its refinancing operations, it does not control longer-term interest rates such as those paid on 10- or 20-year government bonds. Hence, fiscal profligacy by a government in the euro area will tend to push up the cost of borrowing for all members of the currency union.

However interest rates may not be raised enough to discipline properly the high-borrowing government. This is because the markets feel that the other members of the monetary union would not allow the country concerned actually to default, and that if it threatened to do so the other members would probably rush in and buy up its government debt and 'bail out' the country concerned. If the markets believe in this possibility then the debt will not be seen as risky as it otherwise would be and so the interest rates charged to the debtor country on its debt will not be as high as they otherwise might be. The intervention by the ECB, EU finance ministers and the International Monetary Fund (IMF) to bail out the economies of Ireland, Greece and other countries since the financial crisis has been a case in point as ministers struggled to keep EMU together.

The net effect is for that government to pay interest rates on its large stock of debt that are lower because of the implicit belief that it will be bailed out if it has problems servicing the debt, and for all other members of the currency union to pay higher interest rates on their government debt because the government has flooded the financial markets with euro-denominated government bonds. In essence, this is an example of the free rider problem: the government is enjoying the benefits of a fiscal expansion without paying the full costs.

In addition, if that government is using the proceeds of its borrowing to fund a strong fiscal expansion, this may undo or work against the anti-inflationary monetary policy of the ECB by stoking up aggregate demand throughout the whole of the euro area.

In order to circumvent some of these problems, the currency union members can enter into a 'no bail-out' agreement that states that member countries cannot expect other members to come to their rescue if their debt levels become unsustainable, as an attempt to convince the markets to charge

CASE STUDY The Stability and Growth Pact: A Ferocious Dog with No Teeth

The Stability and Growth Pact (SGP) was a set of formal rules by which members of EMU were supposed to be bound in their conduct of national fiscal policy. Its main components were as follows:

- Members should aim to achieve balanced budgets.
- Members with a budget deficit of more than 3 per cent of GDP will be subject to fines that may reach as high as 0.5 per cent of GDP unless the country experiences exceptional circumstances (such as a natural disaster) or a very sharp recession in which GDP declines by 2 per cent or more in a single year.

Clearly, however, if EMU members adhered to the SGP, then it would rule out any free rider problems associated with excessive spending and borrowing in any one member country by forcing members to put a limit on the national government budget. The choice of a maximum budget deficit of no more than 3 per cent of GDP was related to a clause in the 1992 Maastricht Treaty that suggested that a 'prudent' debt-to-GDP ratio should be no more than 60 per cent. This itself was perhaps somewhat arbitrary – although it was very close to the actual debt-to-GDP ratio of Germany in 1992. To see, however, that a 60 per cent ratio of debt to GDP could entail 'prudent' budget deficits of no more than 3 per cent a year, let's do some simple budgetary arithmetic. Suppose a country is enjoying real GDP growth of 3 per cent a year and inflation of 2 per cent a year, so that nominal GDP is growing at the rate of 5 per cent a year. This means that the nominal value of its government debt can grow at a rate of 5 per cent a year and still be sustainable. But if the debt-to-GDP ratio is 60 per cent, this means that debt can increase by 5 per cent of 60 per cent, or 3 per cent of GDP a year while keeping the debt-to-GDP ratio constant. In other words, it can run a budget deficit of 3 per cent of GDP a year.

While, however, there was some logic in setting a maximum budget deficit of 3 per cent a year (given a maximum prudent debt-to-GDP ratio of 60 per cent), it is not clear why the SGP suggested that members should aim for a balanced budget. It should be clear from the budgetary arithmetic just discussed that it is not imprudent for countries to run small budget deficits as long as they are enjoying sustained long-term growth in GDP. The effective straitjacketing of national fiscal policy that the SGP implied may have reflected a desire among the architects of EMU for the ECB to maintain an effective monopoly on demand management, so that its policies could not be countered by national fiscal policies.

The crucial question for the SGP, however, was whether or not the maximum allowable budget deficit would be enough for a country to let its automatic fiscal stabilizers come into play when it goes into recession. This is crucial in a monetary union because member countries will have already given up their right to pursue an independent monetary policy and they cannot use the exchange rate as an instrument of policy.

In practice, the SGP proved to be something of a toothless watchdog. As the euro area experienced sluggish growth in the early years of EMU, several member countries – and in particular France and Germany, two of the largest member countries – found themselves in breach of the SGP excessive deficit criteria. However, both France and Germany managed to persuade other EMU members not to impose fines and, in 2004, the European Commission drew up guidelines for softening the SGP. These guidelines included considering more widely the sustainability of countries' public finances on an individual basis, paying more attention to overall debt burdens and to long-term liabilities such as pensions, rather than to a single year's deficit. The consequences of the financial crisis have made the limitation of the SGP even more clear.



Critics of the SGP suggested that when it came to the test it failed to impose discipline.

profligate spend-and-borrow countries higher interest rates on their debt. In fact, exactly such a no bail-out agreement exists among members of EMU. Unfortunately, however, it seems clear that the no bail-out clause is not credible: the attempts by members to support profligate countries throughout the sovereign debt crisis have confirmed this. If an EMU member were to default on its debt, this would have strong repercussions throughout the euro area as it would lead to the financial markets losing confidence in debt issued by other members and to strong selling of the euro in the foreign currency exchange market. In order to avoid this, EMU members have acted to bail out member countries threatening to default on their debt.

For these reasons, the members of the currency union may wish to impose rules on one another concerning the conduct of national fiscal policies in order to avoid fiscal profligacy by any one member. At the outset of EMU, a set of fiscal rules was indeed drawn up and agreed to by EMU members. This set of rules was known as the Stability and Growth Pact (SGP). The SGP not only laid down strict rules on the maximum permissible budget deficit and debt-to-GDP ratio for EMU members, it also stipulated harsh punishments – fines amounting to as much as 0.5 per cent of GDP – for offenders. In the event, several countries, including France and Germany, breached the SGP but managed to persuade other members not to impose fines.

SELF TEST What is fiscal federalism? • How may it aid in the functioning of a common currency area?

The Fiscal Compact

The problems with the debt crisis faced by a number of European countries in the wake of the financial crisis and the attempts to keep the euro area together led to the negotiation of a treaty which came in to force on 1 January 2013. This has become known as the fiscal compact but its full title is the *Treaty on Stability, Coordination and Governance in the Economic and Monetary Union*. The intention of the fiscal compact is to:

strengthen the economic pillar of the Economic and Monetary Union by adopting a set of rules intended to foster budgetary discipline through a fiscal compact, to strengthen the coordination of economic policies and to improve the governance of the euro area, thereby supporting the achievement of the European Union's objectives for sustainable growth, employment, competitiveness and social cohesion.

The treaty echoes some of the features of the SGP in that it requires the budgets of participating member states to be in balance or in surplus and that the structural deficit does not exceed 0.5 per cent of the country's nominal GDP. A **structural deficit** refers to a situation where the deficit is not dependent on movements in the economic cycle but indicate that a government is 'living beyond its means' – spending what it has not got and contrasts with a **cyclical deficit** where government spending and income is disrupted by deviations in the 'normal' economic cycle.

structural deficit a situation where a government's deficit is not dependent on movements in the economic cycle
cyclical deficit a situation when government spending and income is disrupted by the deviations in the 'normal' economic cycle

Each member country will have a minimum benchmark figure for long-term sustainability which will be reviewed annually. Deviation from the balanced budget rule will be allowed in exceptional circumstances, such as a severe economic downturn but if the member state deviates from the rule, an automatic

correction mechanism which has to be built into the country's legal system, will be triggered. Participant countries will have been expected to have this legal mechanism enshrined in law by 1 January 2014. Member states will be subject to a budgetary and economic partnership programme which will outline detailed structural reforms that will have to be put into place if the deficit rules are breached which will detail how the country intends to remedy its deficit problems. The rules apply to all 25 countries which signed the agreement apart from the Czech Republic and the UK who both opted out of the agreement.

At the time of writing it is by no means certain that these new rules will be any more successful than the SGP. Only months into the new treaty there are concerns that France will breach the rules and a number of economists have noted that the rules do not make sense. Imposing greater fiscal discipline might sound sensible to prevent governments from profligacy and reducing the free rider problem but the consequences are essentially simple – if debt is too high the government concerned must cut spending and raise taxes. If governments are seeking to balance budgets, account must be taken of the effect on the other aspects of the economy.

Recall that savings of households and firms is denoted as S and that in equilibrium, savings equals investment ($S = I$) in a closed economy. If governments have a balanced budgets then tax receipts (T) equal government spending (G): $T = G$. The external current account is given by net exports, $X - M$. If there is an imbalance in private savings and/or government budgets, this shows up in net capital outflow as we saw in Chapter 28. In Chapter 29, we noted the link between the market for loanable funds and the market for foreign currency exchange as $S = I + NX$. This can be rearranged to give $S - I = NX$ and rearranging the national accounting formula gives us $(S - I) + (T - G) = NX$ where $T - G$ is the government deficit. If the government is to balance its budget then the sum of the other elements of the economy must also balance. Bringing these other aspects into balance is not easy.

If the government is forced to cut spending and increase taxes to bring the budget into balance, national income will fall, *ceteris paribus*. An increase in taxes will have the effect of reducing private saving; if saving falls then interest rates rise, investment may decline and the net result is a rise in unemployment. In addition, the reduced investment leads over time to a lower stock of capital. A lower capital stock reduces labour productivity, real wages, and the economy's production of goods and services. The political ramifications of bringing the budget into balance could be severe for governments and lead to political instability which further undermines market confidence.

CONCLUSION

This chapter has developed some of the main issues around common currency areas, focusing in particular on European monetary union. Where there is a high degree of trade among a group of countries, there are benefits to be had from forming a currency union, largely arising from the reduction of transaction costs in international trade and reductions in exchange rate uncertainty. However, there are also costs associated with joining a monetary union, largely associated with the loss of monetary autonomy (member countries are no longer free to set their own interest rates) and the loss of exchange rate movements as a means of achieving macroeconomic adjustment. Any decision to form a currency union must weigh these costs and benefits against one another to see if there is an overall net benefit. Although, in the long run, the loss of exchange rate adjustment and monetary autonomy may have little effect on the equilibrium levels of output and unemployment in the economies involved, there may be substantial short-term economic fluctuations in these macroeconomic variables as a result of joining the currency union. This is particularly the case if there are asymmetric demand shocks impacting on the currency union so that it is impossible to design a 'one-size-fits-all' monetary policy to suit every country. Short-run adjustment will also be long and painful when wages do not adjust very quickly, although this problem may be overcome by labour mobility across the member countries.

A group of countries for which the benefits of monetary union are high and the costs are relatively low is termed an optimum currency area. Even though there is quite a high degree of trade integration among the member countries of the current European monetary union and their economic cycles do seem more or less synchronized and of a similar amplitude (with some exceptions), labour mobility and wage flexibility in Europe are both notoriously low, and integration of euro area financial markets,

although high in the wholesale sector, has so far been disappointing in the retail financial markets. Overall, therefore, the euro area is probably not an optimum currency area. Nevertheless, it is possible that some of these criteria may be endogenous: EMU may lead to increasing economic integration in the euro area that will in turn significantly raise the benefits and reduce the costs to each country of remaining in the monetary union.

IN THE NEWS



A Euro-Budget

The effects of the financial crisis have led to a reassessment of some of the assumptions of the way in which the euro area has been operating. Some economists suggest that each member state ought to contribute funds to create a euro-budget that could be used to help offset asymmetric shocks.

Does the Eurozone Need a Budget?

The Maastricht Treaty created a restrictive European monetary union by keeping fiscal policy national, subject only to a stability pact. Was this a wise choice? Many observers now argue that the Achilles' heel of the euro is that it has no mechanism to ease shocks to individual member states. The euro area, they argue, needs its own budget to provide some automatic insurance to individual countries when they are hit by these asymmetric shocks.

Even proponents of a sizeable 'federal' euro budget admit that permanent shocks (for example the collapse of a major export market) require permanent adjustment in wages and expenditure. But what is lacking is allegedly a mechanism to redistribute funds from countries experiencing temporary booms to those in recession. This, they say, would dampen the normal ups and downs of an economy and make short term shocks easier to absorb. They usually hold up the United States and its redistributive federal system as an example of how this can work.

But a closer look reveals that this doesn't work as well as is widely assumed. It is true that in the US, as in most federal states, the federal budget redistributes income across regions and thus offsets at least part of the inter-regional differences in income. But the inference that redistribution is equivalent to a shock absorber is wrong.

For example, in the US, the federal budget offsets a substantial part of the differences in the level of income per capita across states – generally believed to be between 30 per cent and 40 per cent – because poorer states contribute on average lower income tax and receive higher social security payments. However, this does not imply that these mechanisms also provide an insurance against temporary shocks to individual states. Many of the transfers from the federal government, especially basic social support such as food stamps, etc. change little with the local business cycle. For example, retirees in Florida receive their pensions whether or not the local economy is doing well (as during the real estate boom up to 2007). These pensions do not increase when the local economy enters a

bust, as it did after 2008. This type of transfer payments from the federal government thus does not provide any buffer against local shocks.

On the revenue side the degree to which federal taxes absorb shocks at the state level cannot be very large for the simple reason that the main federal sources of revenues that react to the business cycle, such as federal income tax, accounts for less than 10 per cent of GDP.

This low sensitivity of both federal expenditure and revenues to local business cycle conditions explains why on average only a small fraction – estimated about between 10 and 15 per cent – of any shock to the GDP of any individual state is absorbed via automatic transfers to and from the US federal budget.

A related idea which has come up repeatedly in the European context is to create some European, or euro area, unemployment insurance fund. This idea is very attractive at first sight. But here again the reference to the US experience is misleading: in the US, unemployment insurance is actually organized at the state level. The federal government intervenes only in the case of major nationwide recessions and provides

some supplementary benefits for the long-term unemployed. But this support is given to all states and thus does not provide those most affected with much more support than the others.

Moreover, unemployment benefits are not as important as often assumed. In most countries they amount to only about between 2 and 3 per cent of GDP, even during a major recession. In the US, the supplementary federal expenditure amounted to only about 1 per cent of GDP in recent years. It is thus clear that a euro area unemployment insurance system would never be able to offset major shocks, such as the ones hitting Ireland or Greece where GDP has fallen by more than 10 per cent.

All in all, it is difficult to rest the case for some euro area fiscal shock absorber on the US experience.

If business-cycle shocks were really the key problem, individual member states could first of all 'self-insure' by running a prudent fiscal policy and lower their debt level so that they have the freedom to run temporary deficits in case they face temporary shocks. The

'Fiscal Compact' with its target of approximate balance in cyclically adjusted terms is implicitly based on this idea.

Moreover, those member states that feel most exposed to business cycle shocks could very well agree among themselves to some mutual insurance scheme. There would be no need to make this scheme compulsory for all member states if the aim is only to pool limited risks on an actuarially fair basis. It is of course true that the risk pooling becomes more efficient the more member countries participate (and the less their business cycles are correlated). This implies that if a subgroup of member states starts such a system (in the spirit of what is called in EU legal terms a 'reinforced cooperation') the other member states should be attracted as well. The force of attraction of such a scheme would provide a litmus test of its usefulness.

Proponents of the idea that the euro area needs a fiscal shock-absorbing mechanism might thus do better proposing a concrete mechanism that could be implemented among a subgroup of member states

instead of jumping to the conclusion that all member states be forced to pay a large part of their tax revenues into a euro area budget.

Questions

- 1 What do you think is meant by the statement: 'The Maastricht Treaty created a restrictive European monetary union by keeping fiscal policy national, subject only to a stability pact'?
- 2 Why do you think that the lack of a 'mechanism to ease shocks to individual member states' is described as the 'Achilles' heel of the euro?
- 3 Explain how a 'federal euro budget' would help to reduce the impact of short-term shocks to member economies?
- 4 Why is the argument for a euro-budget based on a comparison with the United States likely to be weak, according to the author of this article?
- 5 Why might a voluntary risk-pooling scheme among interested member states be more likely to succeed, according to the author, than an enforced EU-wide scheme?

There are stark contrasts in economic welfare across Europe. In parts of Italy there are regions suffering very weak economic growth, whereas in parts of Germany the population benefits from the strong German economy, as exemplified by the snow-covered mansions in Heidelberg.



SUMMARY

- A common currency area (a.k.a. currency union or monetary union) is a geographical area through which one currency circulates and is accepted as the medium of exchange.
- The formation of a common currency area can bring significant benefits to the members of the currency union, particularly if there is already a high degree of international trade among them (i.e. a high level of trade integration). This is primarily because of the reductions in transaction costs in trade and the reduction in exchange rate uncertainty.
- There are, however, costs of joining a currency union, namely the loss of independent monetary policy and also of the exchange rate as a means of macroeconomic adjustment. Given a long-run vertical supply curve, the loss of monetary policy and the lack of exchange rate adjustment affect mainly short-run macroeconomic adjustment, however.
- These adjustment costs will be lower the greater is the degree of real wage flexibility, labour mobility and capital market integration across the currency union, and also the less the members of the currency union suffer from asymmetric demand shocks.
- A group of countries with a high level of trade integration, high labour mobility and real wage flexibility, a high level of capital market integration that does not suffer asymmetric demand shocks across the different members of the group, is termed an optimum currency area (OCA). An OCA is most likely to benefit from currency union.
- It is possible that a group of countries may become an OCA after forming a currency union, as having a common currency may enhance further trade integration, thereby helping to synchronize members' economic cycles, and having a single currency may also help foster increased labour mobility and capital market integration.
- While the current euro area displays, overall, a high degree of trade integration and does not appear to be plagued by asymmetric demand shocks, real wage flexibility and labour mobility both appear to be low. And while the introduction of the euro has led to a high degree of euro area financial market integration at the wholesale level, retail financial markets remain nationally segregated. Overall, therefore, the euro area is probably not at present an optimum currency area, although it may eventually become one.
- The problems of adjustment within a currency union that is not an OCA may be alleviated by fiscal federalism – a common fiscal budget and a system of taxes and fiscal transfers across member countries. In practice, however, fiscal federalism may be difficult to implement for political reasons.
- The national fiscal policies of the countries making up a currency union may be subject to a free rider problem, whereby one country issues a large amount of government debt and pays a lower interest rate on it than it might otherwise have paid, but also leads to other member countries having to pay higher interest rates. It is for this reason that a currency union may wish to impose rules on the national fiscal policies of its members.

QUESTIONS FOR REVIEW

- 1 What are the main advantages of forming a currency union? What are the main disadvantages?
 - 2 Are the advantages and disadvantages you have listed in answer to question 1 long run or short run in nature?
 - 3 Is a reduction in price discrimination across countries likely to be an important benefit of forming a currency union? Explain.
 - 4 What are the main characteristics that reduce the costs of a common currency?
 - 5 How would an asymmetric supply-side shock to a country that was a member of a currency union affect the country and the other members of the currency union?
 - 6 What is an optimum currency area (OCA)? List the criteria that an OCA must satisfy.
 - 7 Is EMU an optimum currency area?
 - 8 What is fiscal federalism? How might the problems of macroeconomic adjustment in a currency union be alleviated by fiscal federalism?
 - 9 Why might the members of a currency union wish to impose rules on the conduct of national fiscal policies?
 - 10 What are the main features of the fiscal pact?
-

PROBLEMS AND APPLICATIONS

- 1 Consider two countries that trade heavily with one another – Cornsylvania and Techoland. The national currency of Cornsylvania is the cob, while the Techoland national currency is the byte. The output of Cornsylvania is mainly agricultural, while the output of Techoland is mainly high-technology electronic goods. Suppose that each economy is in a long-run macroeconomic equilibrium.
 - a. Use diagrams to illustrate the state of each economy. Be sure to show aggregate demand, short-run aggregate supply and long-run aggregate supply.
 - b. Now suppose that there is an increase in demand for electronic goods in both countries, and a simultaneous decline in demand for agricultural goods. Use your diagrams to show what happens to output and the price level in the short run in each country. What happens to the unemployment rate in each country?
 - c. Show, using your diagrams, how each country could use monetary policy to reduce the short-run fluctuation in output.
 - d. Show, using your diagrams, how movements in the cob–byte exchange rate could reduce short-run fluctuations in output in each country.
- 2 Suppose Techoland and Cornsylvania form a currency union and adopt the electrocarrot as their common currency. Now suppose again that there is an increase in demand for electronic goods in both countries, and a simultaneous decline in demand for agricultural goods. As president of the central bank for the currency union, would you raise or lower the electrocarrot interest rate, or keep it the same? Explain. (Hint: you are charged with maintaining low and stable inflation across the electrocarrot area.)
- 3 Suppose that Techoland and Cornsylvania decide to engage in fiscal federalism and adopt a common fiscal budget.
 - a. Show, again using aggregate demand/aggregate supply diagrams, how fiscal policy can be used to alleviate the short-run fluctuations generated by the asymmetric demand shock.
 - b. Given the typical lags in the implementation of fiscal policy, would you advise the use of federal fiscal policy to alleviate short-run macroeconomic fluctuations? (Hint: distinguish between automatic stabilizers and discretionary fiscal policy.)
- 4 The United States can be thought of as a non-trivial currency union since, although it is a single country, it encompasses many states that have economies comparable in size to those of some European countries. Given that the USA has had a single currency for 200 years, it may be thought of as a successful currency union. Yet many of the American states produce very different products and services, so that they are likely to be impacted by different kinds of macroeconomic shocks (expansionary and recessionary) over time. For example, Texas produces oil, while Kansas produces agricultural goods. How do you explain the long-term success of the US currency union given this diversity? Are there any lessons or predictions for Europe that can be drawn from the US experience?
- 5 Explain, giving reasons, whether the following statements are true or false.
 - a. ‘A high degree of trade among a group of countries implies that there would be benefits from them adopting a common currency and forming a currency union.’
 - b. ‘A high degree of trade among a group of countries implies that they should definitely adopt a common currency and form a currency union.’
- 6 Do you think that the free rider problem associated with national fiscal policies in a currency union, as we discussed in the text, is likely to be a problem in actual practice? Justify your answer.
- 7 What is the function of the European Commission? What are the other five main institutions of the European Union and what are their respective roles? What are the other important EU bodies and what are their respective roles? (Hint: go to the European Union website: www.europa.eu.int.)
- 8 If the interest rate on Spanish government debt is rising whilst that of Germany is falling:
 - a. What does this tell you about the view of the markets on the two countries?
 - b. If the interest rate is rising, what is happening to the price of bonds for the two countries? Explain.
- 9 In order for a common currency area to work effectively it is argued that on joining, member states need to be at a similar stage in the economic cycle. Why do you think this is the case?
- 10 The fiscal pact would be fine if all countries in the euro area were at the same stage in the economic cycle and it was designed to act as a deterrent for profligacy in the future. As a means of solving the debt crisis in Europe it is doomed to fail. To what extent would you agree with this statement?

37

THE FINANCIAL CRISIS AND SOVEREIGN DEBT

One of the things that makes economics fascinating is the speed with which things can change. In 2005 few people were predicting the calamitous events which would take place over the next three years. Some commentators had warned that the rise in house prices both in the UK and the US would at some point have to suffer a correction but arguments differed about the relative size of that correction and the effects on the economy as a consequence. In a speech in 2004 (<http://www.federalreserve.gov/newsevents/speech/bernanke20070517a.htm>) Ben Bernanke, an acknowledged scholar of the Great Depression, said: 'One of the most striking features of the economic landscape over the past 20 years or so has been a substantial decline in macroeconomic volatility.' He coined this 'The Great Moderation'. Bernanke put down the reduction in volatility to structural change, improved macroeconomic policies, in particular the focus on achieving stable and low inflation rates, and good luck!

The government in the UK was typical of many in the West who were confident of continued macroeconomic stability. In what was his last Budget speech on 21 March 2007, the Chancellor and soon-to-be Prime Minister, Gordon Brown, said:

In this, my eleventh Budget, my report to the country is of rising employment and rising investment; continuing low inflation, and low interest and mortgage rates; and this is a Budget ... built on the foundation of the longest period of economic stability and sustained growth in our country's history ... Our forecast and the consensus of independent forecasts agree that looking ahead to 2008 and 2009 inflation will also be on target... And by holding firm to our commitment to maintain discipline in public sector pay, we will not only secure our 2 per cent inflation target but create the conditions for maintaining the low interest and mortgage rates that since 1997 have been half the 11 per cent average of the previous 20 years. And we will never return to the old boom and bust.

So Mr Deputy Speaker, with consumption forecast to rise in each of the next two years by 2¼ to 22¾ per cent, investment and exports by more than 3 per cent, we expect that next year, in 2008, alongside North America, our growth will again be the highest in the G7 – between 2½ and 3 per cent – with the same rate of growth also in 2009 – under this Government, with stability in this as in every other Budget the foundation, sustained growth year on year.

Within a few months of that speech, the global economy experienced a crisis that has led to one of the deepest and longest-lasting periods of economic downturn and weak growth in history. Given the confidence of the UK Chancellor in early 2007, one might ask questions about whether there were signs of the problems to come and why did such a change in fortunes for so many countries around the world happen so quickly?

BUBBLES AND SPECULATION

One of the areas of disagreement amongst economists was the extent to which the pre-crisis financial landscape was subject to a 'bubble'. An **economic bubble** occurs when prices of assets and securities rise way above their true or fundamental value. Once a bubble bursts, asset and security prices collapse.

The collapse in asset and security prices was a feature of the financial crisis but the origins of the bubble (and some would argue the word is meaningless) lie in the way in which economies changed in the 30 years prior to the crisis.

economic bubble when prices of assets and securities rise way above their true or fundamental value

Deregulation

During the 1980s, the move towards supply-side policies led to a period of deregulation. In particular, rules and deregulation surrounding the activities of financial institutions were abolished or relaxed. In the UK and US in particular, successive governments gradually picked away at regulation during the 1980s and 1990s allowing financial institutions to trade globally and with more freedom to innovate than ever before. Deregulation went hand-in-hand with developments in technology and led to further innovation.

Deregulation meant that the average person in the street had access to credit which was much easier than it had ever been. In the US and UK, this was further enhanced by the rise in house prices which gave many people positive equity in their homes which they could use as security to access credit. What these funds could be used for was also relaxed and so more homeowners were in a position to borrow money to buy cars, holidays, finance weddings and other luxuries as well as carry out home improvements and extensions.

In the US, shocks to the system in the form of the dot.com collapse and the terrorist attacks of 9/11 led to dips in economic activity. In each case the Fed responded by cutting interest rates and flooding the markets with cash. In addition, preparations for a predicted IT disaster at the turn of the millennium (the Y2K threat) had meant that many large financial organizations had invested heavily in technology which meant that they possessed advantages that enabled them to improve productivity – possibly ahead of time if that investment into the Y2K threat had not occurred. Even in the face of concerns about the fragility of consumer spending, profitability in banks rose.

As the search for returns intensified, banks looked to widen the market segments that they were prepared to do business with in the search for new markets and opportunities to improve returns. In addition, the prevalence of relatively low rates of interest meant the prospect of high yield returns became even more attractive, especially given the fact that such returns could be linked to bonuses for staff. Deregulation and global trading meant that capital movements across national boundaries were much easier. Funds could be borrowed from countries with low interest rates (particularly Japan which experienced negative interest rates at times in the early 2000s) and invested in assets in countries with higher yielding currencies.

The Rise in House Prices

The thirst for credit on both sides of the Atlantic, particularly in the form of mortgage lending, has been cited as being a cause of the bubble in house prices. Panel (a) in Figure 37.1 shows UK house prices from 1983–2008. It can be clearly seen that the average house price rose dramatically from the late 1990s to 2007. Panel (b) of Figure 37.1 shows the index of house prices in the US. Between 2000 and 2008, house prices in the US had risen by around 80 per cent.

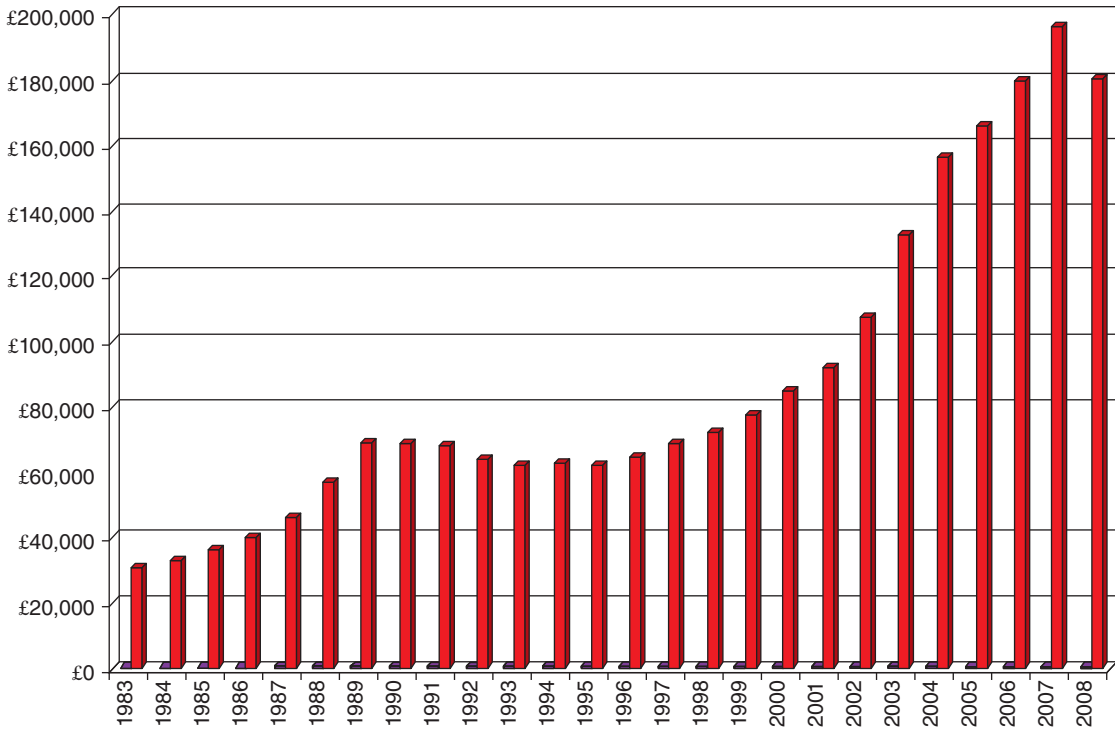
As house prices rise, the difference between the principal borrowed and the value of the house increases – positive equity. An individual may have a mortgage for £250,000 but a property worth £375,000 and thus have positive equity of £125,000. Lenders were not shy in allowing borrowers to exploit positive equity and with the security of a valuable asset (and the expectation that prices would continue to rise) the ease with which homeowners were able to add to their debt increased.

The effects of deregulation, the existence of relatively low interest rates, apparently stable economic conditions and the increase in risk seeking activity amongst financial institutions, all came together to create a rise in house prices fuelled by rising demand and easier access to credit. In the UK, the increase in the number of people being able to access mortgages was the direct result of the willingness of banks and building societies to develop new products to allow those who would previously not have been able to

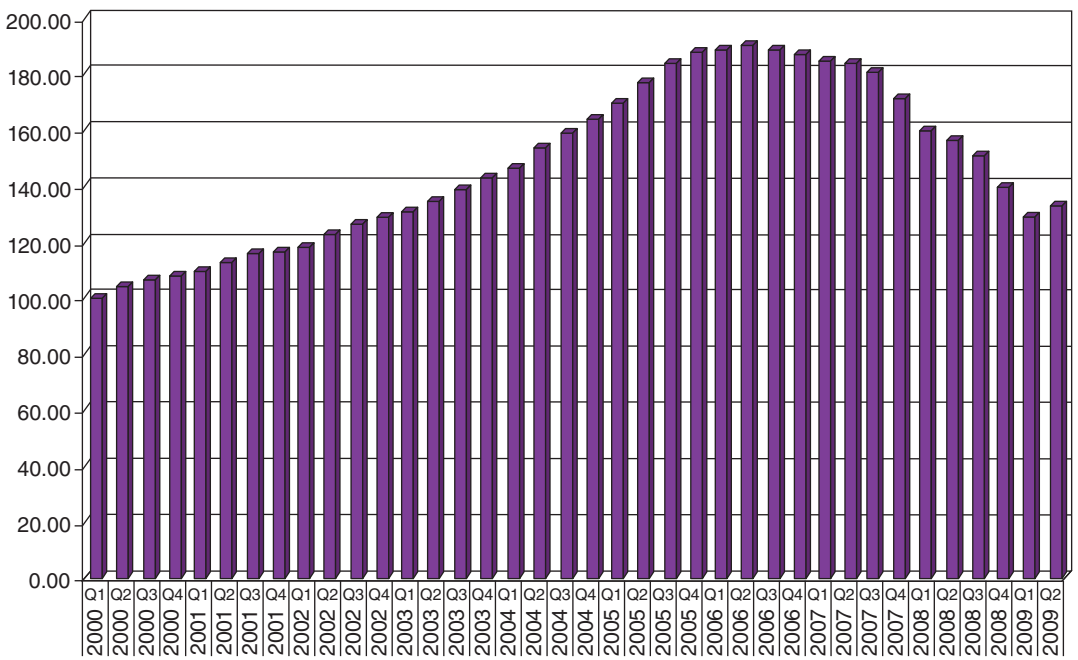
FIGURE 37.1

The Rise in House Prices

Panel (a) shows the rise in the average house price in the UK between 1983 and 2008. The average house price in 1983 was around £30,000 but by 2007 had peaked at around £190,000. Panel (b) shows the Standard & Poors (S&P)/Case–Schiller US National Home Price Index between 2000 and 2009 indicating the sharp increase in house prices in the first years of the noughties.



(a)



(b)

have access to a mortgage, to do so. In the US it was those whose credit histories were very weak who were targeted. As market prices rose the desire to get on the housing ladder further fuelled demand. For those who had managed to access the market in the early 1980s, the amount of positive equity in their properties was such that it allowed some to borrow to fund home extensions, holidays, new cars, home improvements and for a growing number, a second property. This second property might be a retreat near the coast or in the countryside, an apartment in one of the many new developments in Spain or a property bought with the intention of letting it out and using the rent to help pay for the second mortgage.

The Sub-Prime Market

In the United States, extending mortgage opportunities to those not traditionally seen as being part of the market was also part of the way in which banks and other lenders sought to increase their lending. Individuals seen as being an acceptable risk for mortgage lending were known as the prime market; the term is said to have derived from analogy with the best cuts of meat from an animal. It followed that there was a 'sub-prime' group for whom access to mortgages was altogether more difficult. Some of these people had credit histories that were very poor, some did not have jobs, but in an atmosphere of risk seeking and changed priorities this group provided lenders with a market opportunity.

Individuals may have been contacted in the first instance by a broker who outlined the possibilities for the borrower often referred to as 'affordability products'. These products sometimes included so-called 'teaser rates' which offered low interest rates for an initial period, often two years, and in some cases zero interest rates for the first two years. Interest rates after the first two years could rise significantly but the attraction for many prospective house buyers was that the chance to access mortgage funds was being made available to them for the first time. Typical of these types of loan were 2/28 or 3/27 loans. The 2 and 3 indicated the number of years of the teaser rate and the remainder the period over which a variable rate would apply.

Brokers would then look to sell the mortgage to a bank or other mortgage provider and received a commission for doing so. Under 'normal' circumstances the bank would not only have acquired a liability but also an asset (the payments received by the borrower on the mortgage). The mortgage would generate a stream of cash flows in the form of the payment of the mortgage principal and interest over a period of years. Such an asset could be valuable in the new deregulated banking environment and led to the securitization of these assets which we covered in Chapter 26.

The growth of sub-prime mortgages began to take an increasing share of the total mortgage market in the US and the risk involved in terms of the loan to value of property also rose from around 50 per cent to over 80 per cent by 2005–2006. Sub-prime loans accounted for around 35 per cent of all mortgages issued in 2004. It was not only the poor who were taking advantage of the sub-prime market, however. Existing homeowners took advantage of the positive equity in their properties to re-mortgage or buy second homes. Speculation in the housing market was not confined to just those in the financial markets that could benefit from new financial instruments that could be traded for profit. Some saw opportunities to enter the housing market, borrow funds to buy property and almost immediately put it back onto the market – a phenomenon known as 'flipping'. The hope was that rapidly rising house prices would mean that the property could be sold quickly and a profit realized.

The macroeconomic background to this was a sharp cut in interest rates by the Fed in response to the collapse of the dot.com boom and also in response to 9/11. A low interest rate environment seemed to insulate sub-prime mortgage holders from the extent of the debt which they had taken on. From 2003 onwards, interest rates stayed at 1 per cent and for 31 consecutive months the inflation-adjusted interest rate was effectively negative. Meanwhile the rise in house prices further contributed to economic growth as employment in housing related business expanded to meet the growth in activity. The housing market was responsible for creating over 788,000 jobs in the US between 2001 and 2005 – some 40 per cent of the total increase in employment.

With the benefit of hindsight it might seem rather foolish for anyone, let alone a bank, to have contemplated lending money to people with bad credit histories and low incomes. Surely banks knew the risk they were taking? Part of the answer is that they *thought* they did and whilst risk seeking might have been a more dominant behavioural trait than risk aversion, strategies had been put in place to limit the risks of the banks and other lenders to the exposure of lending to the sub-prime market. As we have seen in

CASE STUDY The South Sea Bubble

In 1720, Isaac Newton is quoted as saying 'I can calculate the motions of the heavenly bodies but not the madness of people'. Newton had made a profit of £7,000 (about £800,000 or €937,500 in today's values) selling shares of the South Sea Company in April of that year. Along with many other people, however, Newton believed that the value of the company would continue to rise and he purchased more shares as the price of the company continued to rise. Newton ended up losing £20,000 (about £2.3 million or €2.7 million).

The South Sea Company was formed in 1711. It entered into an agreement to assume part of the debt of the government, which at the time was financing the War of Spanish Succession, in return for a monopoly on the trade to Spanish colonies in South America. The Treaty of Utrecht was signed in 1713 to end the war. Its terms were not favourable to the South Sea Company as it limited trade opportunities to the Spanish colonies. A further debt conversion was announced in 1719 where the South Sea Company became just one of three corporations, including the Bank of England, which owned around 36 per cent of the total national debt. Despite the fact that the company had no real evidence of trading success, the directors of the company continued to make claims to shareholders of the wealth and riches that trade promised, not least large quantities of gold and silver just waiting to be brought back to Europe.

In January 1720, the South Sea Company's share price was £128. The claims by the Company pushed up the share price and more dabbling in debt conversion with Parliament saw its price rise to £330 by the end of March. Newton sold his shares in April; by May the price had risen further to £550. In June Parliament passed the Bubble Act which was introduced by the South Sea Company and required all joint-stock companies to acquire a royal charter. Part of the reason for the legislation was to help control the explosion of companies entering the market all making claims for their ventures which were sending prices rising causing mini-bubbles. Cynics would argue that the South Sea Company also used the legislation to manage the growing threat of competition it faced. Having received its royal charter, shareholders took this as being a sign that the company's claims were sound and demand for its shares continued to rise. By the end of June the share price had risen to £1,050.

For some reason (and this perhaps is where Newton's lament over the madness of people is most pertinent), some people began to sell their shares and the share price began to fall in July. For those that had come in near the top of the market (like Newton's second purchase) the imperative to sell became more urgent. The price continued to fall quickly and by September stood at around £175. The collapse affected thousands of people and a number of institutions. Subsequent investigations into the collapse revealed bribery and corruption practices and prosecutions of company officials and members of the government.

The South Sea Bubble was an example of how asset prices can rise way above their true market value. It seems that history can repeat itself and in many ways, the issues that existed then (a lack of knowledge by participants in the market) is similar to today. The complexities of how markets operate may be far

greater and the technologies far more sophisticated, but the success of markets relies to a large extent on participants having good information on which to base their decisions. No matter how intelligent an individual is, it seems, if they do not have access to all available information then poor decisions can be made as Newton found out to his cost.



The South Sea Scheme a caricature by William Hogarth, an 18th-century sociocritical English painter and graphic artist.

earlier chapters, new products which securitized the loans that were being made along with the development of credit default swaps (see Chapter 24) to help insure against the risk of default, changed the nature of banking as we outlined in Chapter 26.

In the case of sub-prime loans the individual taking out the mortgage clearly has some credit risk by definition. Individually such a loan has a credit risk which is likely to be much higher compared to a loan made to a corporation. It would be unlikely that the risk for such a debt could be passed on to someone else. However, bundling up a collection of such debts makes the proposition somewhat different. Putting together a collection of loans and assessing the risk of default was subject to statistical analysis. A proportion of the collection is likely to be subject to default; a mathematician or statistician can provide detailed analysis of the extent of the risk involved (which is why such skills are highly valued in financial markets). Collectively, therefore, these loans can be offered as a far more attractive proposition to a potential group of investors looking to increase their earnings.

The incentives for those involved in banks and other financial institutions were to assume a risk-seeking approach, therefore. The demand by shareholders for better returns, the existence of a bonus culture which could provide some individuals with millions each year, the desire by individuals to get on the property ladder and, some argue, the lack of rigour in regulation and oversight by the central banks combined to provide a situation where the bubble could feed on itself and generate ever greater lending and the search for ever more attractive investments.

SELF TEST Why did financial institutions increase lending to the sub-prime market? Was such lending always doomed to failure?

The Bubble Bursts

The use of credit default swaps (CDS) represent sound business principles when the risk of default is very low which was the case when these were first developed in the late 1990s and when the bonds being insured were corporate bonds with a low risk attached to them. The expansion of CDS presented a different challenge to the financial markets when the collapse came. It must be remembered that not only were these contracts taken out in relation to bonds linked to sub-prime loans but the trade of these contracts might mean that banks and other financial institutions might be buying the contract whilst not possessing the asset (the bond). Contracts could also be bought and sold as part of speculative deals which means the web of financial interdependence grows – all based, remember, on a collection of sub-prime mortgages. Traders might, therefore, deliberately buy CDS on the expectation that the bond associated with it would fail. If it did then the trader stands to gain a return in the form of the bond value which has been insured or the collateral which has been set aside. The market progressed from being one which insured bond holders against risk to one where accurate speculation could generate large returns for dealers. Those buying these derivatives (so called because their value is based on something else) may have little if any idea of the strength of the underlying assets associated with their trade.

The CDS market was also not subject to any form of regulation. It is referred to as an 'over-the-counter' market, a private financial transaction between two parties. There is no official record, therefore, of who has made a trade and whether the seller can actually meet the obligations it is entering into. The precise value of the CDS market by 2007 is difficult to pinpoint for this reason but it has been estimated at between \$45 and \$60 trillion. To put this into some sort of perspective, \$60 trillion is more than global GDP which is estimated at around \$55 trillion! The number of CDS trades, the number of derivative trades and precisely who owns all these trades are not easy to identify. A number of major banks had exposure to CDS including Lehman Bros, Icelandic banks, Barclays and RBS, and in addition insurance companies like AIG were also heavily committed.

In 2006–2007 the prevailing concern among central banks was over the development of inflationary pressure. In the UK, for example, inflation as measured by the Consumer Prices Index rose to a high

of 5.2 per cent by September 2008 but the Bank of England had started to increase interest rates in June 2006 and a series of quarter-point rises saw Base Rate increase to 5.75 per cent by July 2007. In the US, the Fed raised interest rates for 16 consecutive months to June 2006 and the Federal Funds Rate reached 5.25 per cent by the autumn of 2006. As interest rates rose, borrowers, especially those on sub-prime mortgages, began to feel the pressure and reports of the number of defaults on sub-prime loans began to rise.

Once mortgage defaults began to rise to alarming numbers the number of financial institutions affected and their exposure became clearer. The whole edifice was based on the expectation that the underlying assets would continue to generate the income stream over time – in other words, the mortgage holders would continue to pay their monthly mortgage repayments. How did these mortgage defaults begin to happen? We have already seen how teaser rates tempted many to take on the burden of a mortgage but this was not necessarily accompanied by a full understanding of the financial commitments being undertaken. After the teaser rate period finished the mortgages had to be restructured based on a new rate of interest. The new rate of interest not only reflected the need to claw back some of the interest not paid in the initial teaser rate period but also the fact that these borrowers were higher risk.

As mortgage rates rose, borrowers found it increasingly difficult to meet their monthly payments. In some cases the difference between initial monthly payments and restructured payments based on new interest rates was significant; many homeowners found their payments more than doubling. The sub-prime market contained a large number of people who were least able to cope with significant changes in monthly payments but in addition those that had taken out sub-prime mortgages for second homes and the buy-to-let market also found themselves stretched.

Faced with payments they could not afford, borrowers looked to either sell their property or face the prospect of foreclosure – the point where the mortgage lender seizes the property. As the number of foreclosures increased, house prices fell. As house prices fell, more and more people were caught in negative equity. The option of selling the house to get out of the debt was not an option for many; they would still be left with a large debt which they had no hope of paying off. The situation was particularly bad in certain areas of the US. The house price bubble had been focused on particular areas in Las Vegas, Los Angeles and Miami. These were areas that were hit worse by the rise in foreclosures. As the number of foreclosures rose the supply of housing increased and, true to basic economic principles of supply and demand, the price of houses fell further.

Note that foreclosure means seizing the property – to get to this stage a number of other routes and methods to help recover the debt will have been taken so this really represents the last straw in the process. Once foreclosure has occurred then the bank or mortgage lender acquires the property as an asset against the loan but then has to incur the cost of managing the property – maintaining it, paying property taxes and insurance, for example, as well as the cost of trying to sell it to realize the cash. In times of falling house prices the cash realized on the sale may be far less than the original value of the asset. As a result banks have to write down the value of their assets.

As the sub-prime market collapsed, the exposure to bad debt started to become more obvious and a number of banks reported significant write-downs and losses. Confidence in the banking system, so important to its functioning, began to fall. Banks were not sure of their own exposure to these bad debts (referred to as toxic debt) and so were also unsure about the extent of other banks exposure. Interbank lending began to become much tighter as banks were unwilling to lend to each other and also faced the task of trying to shore up their own balance sheets. Interbank lending effectively ground to a halt. (See Chapter 26 for an outline of how the process unwound.)

The seizing up of credit markets had an effect around the world. Banks in Iceland failed; in eastern Europe the Ukraine applied for a \$16.5 billion loan from the IMF whilst Hungary looked for \$10 billion, this after receiving a €5 billion loan from the European Central Bank and raising interest rates by 3 per cent to 11.5 per cent. Turkey, Belarus and Serbia also sought financial assistance as credit dried up and banks around the world found they had increasing difficulties in meeting their liabilities. Belarus looked for a \$2 billion loan from the IMF and Hungary. In the Ukraine, the currency, the hryvnia, fell by 20 per cent and around 80 per cent was wiped off stock market values. The Ukraine central bank was forced to use reserves to try to buy hryvnia in an effort to support the currency. The Russian stock

market fell by around 66 per cent between May and October 2008. Put simply, most of these banks had lent too much money, had been exposed to the debts related to purchases of CDS and could not meet payments which were falling due. Without financial assistance these institutions could have collapsed and financial systems in these and associated countries were grinding to a halt.

The Collapse of Lehman Lehman Bros had been a bank at the forefront of the sub-prime and CDS market. It had built its business on leveraging – borrowing heavily to finance its activities. Over the weekend of 13 and 14 September 2008, last-ditch talks were convened to try and find a way of rescuing Lehman Bros. The debts at Lehman and the extent of their exposure to sub-prime and CDS were too great, however, and Lehman Bros filed for Chapter 11 bankruptcy protection on 15 September. When Lehman collapsed it led to billions of dollars of claims on Lehman CDS – payments which had to be made by those who sold the protection. As a means of settling these claims, Lehman's bonds were sold via auction in early October 2008. The value of the bonds was set by the auction at around 8 cents to the dollar. So a \$10 million bond would only be worth \$800,000 meaning that protection sellers would have to make up the remaining \$9,200,000 to meet the CDS obligation. With claims estimated at anything between \$400 and \$600 billion this put a massive strain on the financial system and many banks who had sold CDS on Lehman were now exposed and in danger of failing themselves, which put further pressure on the system. Basically, banks found that they had to meet payment obligations and did not have the funds available in reserves to do so. Banks who were struggling to meet their obligations may have looked to the interbank market to raise funds which we have already seen, was almost at a standstill. Banks around the world found themselves in this difficult position as the extent of the web of securitization began to become clearer. As the problems surrounding securitization unwound, banks had to take some of the assets back onto their balance sheet. This then meant that they had to reinstate sufficient reserves to support these liabilities and this further reduced their willingness to lend. Instead the focus shifted to trying to restore balance sheets by taking in deposits but not lending.

The drip feed of news from banks, large and small, regarding the perilous position they were in sparked a drop in confidence in the whole system. If Lehman could fail what about other banks? Banks that had to make payments would either have to deplete their capital and thus become seriously overstretched and thus at greater risk, or take on the liabilities of the special purpose vehicles (see Chapter 26) and take them onto their balance sheet further depleting their positions. As their positions worsened their credit ratings would fall; if this happened and they held CDS seller positions this would further increase the collateral calls on them and put them under further pressure.

The collapse of Lehman was seen as a spectacular event because it showed that central banks were not prepared to step in to help any bank that got into trouble. Bear Stearns and Merrill Lynch had survived by being bought by J.P. Morgan and Bank of America and the US government had stepped in to bail out the mortgage lenders, Fannie Mae and Freddie Mac. The collapse of Lehman Bros was closely followed by it becoming clear that insurance group AIG was in trouble. AIG had been an active seller of CDS and the collapse of Lehman had left it seriously exposed. The US government stepped in to support this company because so much of the rest of the economy and financial institutions were implicated if AIG had been allowed to go to the wall. In the UK, Northern Rock, RBS and HBOS were all subject to support of various kinds including the speeding up of permission for Lloyds TSB to take over HBOS before it collapsed. Some have suggested, however, that Lehman may have been a sacrificial lamb at the altar of moral hazard.

The Path to Global Recession

The problems in financial markets began to translate themselves to the real economy fairly quickly. As the housing market slumped, many people involved in real estate were affected. The number of jobs created in the US that were linked to the house price boom began to be reversed. As confidence fell, economic activity declined across many parts of the world. The effect of the tight credit market helped exacerbate the depth of recession in many countries. Businesses, especially small businesses, rely on cash flow for

their survival. Many need flexibility and understanding from their banks when faced with shocks to their cash flow and the tight credit markets meant that many businesses found banks unwilling to lend to them and to extend overdrafts. If the bank was prepared to lend, the interest rate attached to the loan was often prohibitive. Faced with cash flow problems and an inability to get the finance to support them many smaller businesses began to close. Larger businesses were not immune from the problems. The motor vehicle industry, in particular, suffered significant falls in sales. In January 2008, monthly production of vehicles in the US was 817,767; by February 2009 this had fallen to 388,267 – a fall of 52.2 per cent in just over a year. The fall in production affected component suppliers and the negative multiplier effect worked its way through the economy.

As demand fell, businesses closed and unemployment rose. In the US, the unemployment rate rose from 6.2 per cent in August 2008 to 9.7 per cent a year later (US Bureau of Labor Statistics), the unemployment rate in the UK rose by 2.4 per cent over the same period and unemployment across the EU reached 9.7 per cent in 2010–2011. The increase in unemployment impacts on business confidence. High street sales in many countries slowed, spending on goods and services with a relatively high income elasticity of demand also fell and as this happened business saw revenues falling with the resulting pressure on cash flow. Many businesses cut back on stocks and this further affected businesses along the supply chain. The Boston Consulting Group (BCG) produced its *Global Wealth Report* in September 2009 which looked at the asset management industry and estimated that global wealth had fallen by 11.7 per cent in 2008 to around \$92 trillion (about €72 trillion or £60.5 trillion).

The Role of Central Banks

As the financial crisis unravelled, focus centred on the response of central banks around the world. The role of monetary policy in helping to alleviate the worst effects of the financial crisis and the resulting recession is seen as important and has since come under scrutiny with regard to the effectiveness of the action of key policy makers. The markets were looking for three key aspects of policy response from central banks: speed of intervention, innovation and coordination. There were aspects of each of these that have occurred since the crisis developed but there are also important differences in the degree to which each of the main central banks responded, partly because of the different levels of flexibility that each enjoyed.

The early signs of crisis emerged in 2007 when rumours of rising levels of default from sub-prime mortgages began to increase. In July, Standard & Poors and Moody's downgraded ratings on bonds backed by sub-prime mortgages. In late July a German bank, IKB, said that it was in financial trouble as a result of its exposure to sub-prime and was swiftly taken over by German government-owned bank, KfW (Kreditanstalt für Wiederaufbau). By early August the ramifications of the downgrading of funds related to sub-prime were starting to unravel and, on 3 August, Bear Stearns had to contact its shareholders following the collapse of two hedge funds that it managed and the subsequent fall in its share price. On 9 August, French bank BNP Paribas announced that it was ceasing trading three of its funds and with that credit markets effectively froze. Interbank lending ground to a halt as banks began to recognize that their exposure to sub-prime could be extensive.

The freezing of credit markets effectively marked the start of intervention by central banks around the world. The ECB was one of the first to react on 9 August with the then President, Jean-Claude Trichet, authorizing an injection of €95 billion into the financial markets to help ease overnight liquidity problems. On the following day a further €61 billion was authorized. Soon after the ECB move the Fed announced that it would inject \$38 billion to help ease liquidity; the Bank of England, however, did not follow suit. On 10 August, the Bank was involved in the growing problems at Northern Rock which had expanded rapidly in previous years but now found its expansion based on leverage unsustainable and, as credit markets froze, it was struggling to continue. In the coming weeks the extent of the problems at Northern Rock became public and a run on the bank ensued. Queues formed outside Northern Rock branches as worried account holders looked to withdraw their money. Moves to calm the situation did not seem to have too much effect and on 17 September the government, in consultation with the Bank, agreed to guarantee all existing Northern Rock deposits.

As banks around the world began to falter over the coming months; the question of which bank should be rescued and which left to fail had to be considered by central bank leaders. The Bank of England had issued warnings that it was concerned investors were not pricing risk appropriately. In the US, Ben Bernanke had

made a number of statements attempting to calm the fears over the extent to which sub-prime would affect the real economy. In May 2007 he had given a speech at a conference in Chicago and said:

We believe the effect of the troubles in the subprime sector on the broader housing market will likely be limited, and we do not expect significant spillovers from the subprime market to the rest of the economy or to the financial system. (<http://www.federalreserve.gov/newsevents/speech/bernanke20070517a.htm> accessed 27 October 2009)

He reiterated this view a month later when he said:

At this point, the troubles in the subprime sector seem unlikely to seriously spill over to the broader economy or the financial system. (<http://www.federalreserve.gov/newsevents/speech/Bernanke20070605a.htm> accessed 27 October 2009)

Whether this was an attempt to exert some psychological influence over decision-makers or a lack of information or understanding of the situation is open to debate. In hindsight, it is easy to isolate quotes and point fingers at central banks. The reluctance of the Bank of England to inject funds into the markets was partly seen as the concern with moral hazard but also the subjugating of financial stability to a lesser role. However, the Bank of England, at that time, had limited tools available to it to deal with the events that were unfolding in comparison to other central banks. It was not until the passing of the Banking Act in 2009 that the Bank of England was given additional powers and responsibilities that enabled it to improve its control of financial stability.

In the event, the Bank of England eventually announced that it would provide financial help to institutions that needed overnight funds but that any such borrowing would incur a penalty rate. By April 2008 it had gained the authority to lend to banks against mortgage debt – something that the ECB could do which the Bank could not. The Bank was given authority to issue short-dated UK government bonds in exchange for mortgage securities. However, at the end of August 2007 it emerged that Barclays had asked the Bank for a loan of £1.6 billion and the news, which would not normally have made front page headlines, was interpreted as being another sign that a major bank was in trouble. Barclays needed the funds because of a malfunction in computer systems related to clearing and had tried borrowing from the wholesale markets. The Bank then removed the penalty rate, since it seemed that far from calming nerves it merely served to increase the sense of panic.

As the crisis began to gather momentum, monetary policy around the world was eased. The Fed cut the repo rate by 0.75 per cent in January 2008; the ECB held its rate at 4.0 per cent. The Bank had cut rates by a quarter-point in February and April 2008 but by May that year the Fed had cut rates seven times in eight months with the Fed Funds Rate standing at 2.0 per cent. As 2008 progressed the financial situation did show clear signs that it was ‘spilling over’ to the economy and output levels began to contract in economies across the globe. The problems were such that, in October, seven major economies, the UK, US, China, the EU, Canada, United Arab Emirates (UAE) and Sweden, announced a coordinated 0.5 per cent cut in interest rates. By then end of October the Fed cut again to 1 per cent.

In the US, discussions were taking place to set up the Troubled Asset Relief Plan (TARP), a \$700 billion plan to support the banking system. In November the Fed announced a further \$800 billion support fund. In that same month the Bank reduced rates by 1.5 per cent – the largest single change in rates since it was given independence in 1997. By March 2009 the MPC had cut interest rates in the UK to 0.5 per cent, the lowest since the Bank of England was established in 1694. The Fed cut its rates to a target of between 0 and 0.25 per cent and the ECB had cut rates to 1 per cent.

The scale of the intervention by central banks and the various fiscal stimulus programmes announced by governments made it clear that the crisis was serious. The ‘contagion’ from the sub-prime fall-out and the collapse of banks around the world accompanied by the alarming declines in output meant that global recession was now the real threat and not the prospects for inflation, which remained stubbornly high in countries like the UK.

There are those who believe that, given the circumstances, central banks acted according to the three key aspects of policy response outlined above. The speed with which central banks intervened and the extent of these interventions did differ, largely because of technical as well as ideological differences, but many see their role as having been decisive in exceptional circumstances. When inflationary pressures eased then central banks did relax monetary policy and acted in a coordinated way in October 2008 to

bring down interest rates. Central banks have had to be flexible and innovative in dealing with the issues that were highly unusual. As with the Bank of England, central banks have expanded their roles and assumed new powers and responsibilities and introduced new tools and instruments. Two such examples have been the growth of bilateral swap agreements between central banks of different countries where local currencies can be swapped against the US dollar to enable trade to be financed and liquidity to be eased and the use of quantitative easing.

SELF TEST Use the concept of the reverse multiplier to explain how the financial crisis led to a global economic downturn.

Lessons Learned?

The Organization for Economic Cooperation and Development (OECD) trace the development of the financial crisis through a series of four main stages.

1. A drive on the part of politicians to widen access to home ownership to the poor. Changes to regulations to facilitate this drive led to a growth in lending that was not prudent. This political impetus was evident through successive US Presidents from Clinton onwards and in the UK via New Labour. As a result the expansion of sub-prime lending in the US and cases of banks in the UK lending at 125 per cent of the value of homes, wider access to credit cards and the build-up of debt and limited background checks for credit worthiness were not acted upon by regulators with sufficient vigour.
2. Changes in regulatory structures, particularly in the US, which allowed entry of new businesses into the mortgage market.
3. Basel II regulations which created incentives and the conditions for banks to develop off-balance sheet entities. (Basel II refers to a framework of regulations developed through discussions at the Bank for International Settlements (BIS).)
4. Changes in policy by national regulatory authorities such as the SEC in the US and the FSA in the UK which allowed banks to change leverage ratios from around 15:1 to 40:1. (Leverage ratios, such as debt-to-equity ratios, measure the proportion of debt to equity.)

The OECD has been critical of national regulatory bodies suggesting that there were weaknesses in the way banks were regulated and that regulatory frameworks had not only failed to prevent the financial crisis but had been culpable in contributing to it. It identified a number of key causes which include:

- the bonus culture
- credit ratings agencies
- failures in corporate governance
- poor risk management strategies and understanding.

The International Monetary Fund (IMF) has broadly concurred with the OECD in its analysis of the key issues. It suggests that financial institutions and investors were both too bullish on asset prices and risk. The low interest rate environment and the extent of financial innovation (encouraged by changes in regulation) allowed excessive leverage to be carried out which increased the web of interconnectedness of financial products but at the same time rendered the inherent risks more opaque. It highlighted the lack of coordination between regulatory bodies and the legal constraints which prevented information sharing to be more widespread thus helping authorities to be able to understand what was going on.

This fragmented approach mean that there were differences in the way in which national regulatory bodies dealt with bank failures and insolvency when in many cases these banks had a global presence not reflected by a global coordinated response by the regulators. The actions that were taken have been described as being 'piecemeal' and 'uncoordinated' which not only led to a weakening of the impact of the policy response but also to market distortion. It also pointed to the lack of appropriate tools available to some central banks to provide the necessary liquidity support in times of crisis, as outlined above.

Other criticisms of the regulatory regimes in place throughout the world highlight the fact that the rules that are in place may not be appropriate to deal with the pace of change in financial markets, that

regulators spend too much time ‘ticking boxes’ rather than identifying poor practice and intervening. There have been accusations that the existence of rules means that regulators are able to hide behind them and shift blame. For example, there were a number of subsidiaries of the three Icelandic banks which failed (Kaupthing, Glitnir and Landsbanki) operating in the UK but the FSA argued that these were, technically, outside its jurisdiction. In the US it has been estimated that the total assets of entities which were outside the banking system and the scope of regulation, but which act like banks, is as big as the ‘official’ banking system itself – a value of around \$10 trillion in late 2007.

Part of the reason for these problems stems from internal problems of the regulators themselves. To have a high level of understanding of the financial system to be able to regulate it effectively, employees of the regulators have to be highly experienced and knowledgeable about the system itself. Over-regulation, it has been argued, was partly responsible for creating the incentives for financial innovation to generate improved returns. These new products were complex and highly interconnected to the extent that regulators and government financial departments had insufficient understanding of them. It has been argued that a number of senior executives in the banks themselves did not fully understand the complexity of securitization models, lacked the skills in asset valuation techniques and risk models and were unaware of the extent to which ‘tail losses’ (the extremes of the normal distribution) could impact on their operations. One reason for this was that such models were based on statistics from the ‘good times’ and had not been ‘tested’ by a downturn. If those at the forefront of such operations did not understand what they were dealing with is it possible to expect those working at the FSA, for example, to do so?

Recruitment of the expertise and skills necessary to staff regulatory bodies effectively is a further issue. Why work for the FSA, for example, for a salary of £116,000 when the skills possessed by individuals of the calibre to work in the FSA could be sold to other sectors of the industry for many millions? Without the resources to do the job properly therefore, regulators will always be hampered.

THE SOVEREIGN DEBT CRISIS

One of the consequences of the financial crisis has been the focus on the problems faced by a number of countries in Europe in managing debt. Budget deficits are the difference between the amount of tax revenue and government spending. Deficits are financed by government borrowing and as borrowing increases the overall debt increases. Governments have to manage this debt by ensuring that they have enough funds to be able to pay back government bonds when they reach maturity and also to pay the interest on the bonds which are outstanding. Lending to governments has invariably been seen as being relatively risk-free; in the wake of the financial crisis it became clear that for some governments, this was not the case.

The macroeconomic shock created by the financial crisis led to a global slowdown in economic activity. When countries experience a prolonged period of economic slowdown, tax revenues decline but spending on welfare systems increases. Whilst automatic stabilizers moderate the effects as we have seen, in a severe recessionary period these automatic stabilizers can be diluted. In this environment, governments need to increase borrowing but the confidence of markets in the ability of some governments to be able to pay back their borrowing means that some countries can be brought to the verge of bankruptcy. In other words, they cannot raise enough funds to be able to meet debt and everyday public spending obligations.

The Greek Debt Crisis

The country which has come to epitomize the sovereign debt crisis is Greece. The problems faced in Greece have been replicated to some extent in Ireland, Spain, Italy and Portugal. In 2009, Greece announced that its deficit would reach almost 13 per cent of GDP, double the amount it had forecast a year previously. In early 2010 the country’s debt was reported as €300 billion; more than the value of its gross domestic product (GDP). In fact its debt was 115.1 per cent of its GDP (remember that the original Stability and Growth Pact rules laid down that debt should not exceed 60 per cent of GDP). The budget deficit rose to nearly 14 per cent – well beyond the 3 per cent required as the terms of membership of the European

Monetary Union. Other members of the EU, notably Germany, accused Greece of being 'profligate' and living beyond its means on the back of its membership of the euro area.

Greece had to raise around €50 billion in 2010 to meet its debt obligations. On 19 May 2010 it needed to pay €8.5 billion to bondholders. On 25 January 2010 the Greek government went to the markets to borrow money in the form of its first bond issue of the year. In the event, the issue was oversubscribed as investors sought to pick up the bonds. Why? A simple reason, the interest on the bonds had to be high to persuade investors to take the risk. It was reported that the issue was valued at around €5 billion with a coupon of 6.12 per cent. Given that interest rates around the world at that time were at record low levels this was high. The spread between the interest Greece had to pay to borrow compared with the Germans was almost 4.0 per cent. The wider the spread on different financial instruments (the difference in the coupon of similar bonds in this instance), the more the market is factoring in the risk of default on the bonds considered more risky – the Greek bonds in this case.

In February and March of 2010, the Greek government had tried to take a stance on public spending, proposing major cuts in jobs, pensions, wages and services. Greek workers took to the streets in protest. It seemed as though the government would find it difficult to implement the sort of fiscal cuts necessary to build confidence with the markets and on the back of the announcements about the increasing size of its debt and the deficit, by April it seemed likely that it would find it difficult to raise further money to meet its obligations. Ratings agencies steadily cut the country's ratings until by late April, Greek debt was officially classed as 'junk'.

The spread between Greek and German bonds continued to rise into April reaching 19 per cent on two-year bonds and 11.3 per cent on ten-year bonds. The nervousness on the financial markets over Greece's debt problems began to spread to other European countries. It became a real possibility that Greece would default and be forced to leave the euro. This fear led to a sharp drop in the value of the euro – if there is an expectation that the price of something is going to fall then there is a possibility of making some money and that is exactly what happened in February 2010. Data from the Chicago Mercantile Exchange (CME) showed that short positions against the euro from hedge funds and traders rose sharply. Traders took positions on the expectation that the euro would fall in value. Traders taking out contracts that the price of the euro would fall could exercise these contracts and make a profit if and when the euro fell in value. Data from the CME showed that over 40,000 contracts had been taken out against the euro with a total value of around \$8 billion.

The crisis began to gather momentum and the markets looked to other EU governments to organize a bail-out. On 10 April 2010, the finance ministers of the euro area announced an agreement on a package of loans to Greece totalling €30 billion. Greece said that it did not intend to use the loans and instead rely on its 'austerity measures'. The extent to which Greece could deliver on these austerity measures was something the financial markets were not convinced about.

A week later the Greek Prime Minister, George Papandreou, finally bowed to what many thought was the inevitable and announced that Greece would take advantage of the emergency loans. Negotiations took place with the EU and the IMF on the structure of these loans, which were predicted to rise to €100 billion.

Whilst it was generally accepted that Greece needed the financial support there were questions raised about the terms under which the loans were to be given. The German government was financing a proportion of the loans with €8.4 billion being spoken of as a possible figure. In order to appease German taxpayers who were not supportive of bailing out the 'profligate Greeks', Chancellor Angela Merkel, under pressure from German taxpayers to take a strong line, insisted on very strict terms and a condition that the Greek government make significant cuts to public spending. The argument that Greece has exploited the free rider problem within a common currency area as we outlined in the previous chapter seemed to be strong. For many Greeks, there was a feeling that their future was being dictated by outsiders, most notably the Germans. The Germans argued that it was unfair that its taxpayers should have to suffer to bail out a country which had clearly not played by the euro rules. On May Day, traditional worker protests had a new focus and the extent and severity of the violence which broke out shocked many. Austerity was clearly not going to be easy to implement and the strength of feeling against the Germans was clear. The Greek government was being squeezed from all sides. The financial markets were nervous that the problems in Greece would spread to other high debt countries (so-called 'contagion effect') and there were fears that the crisis could tip Europe back into recession.

The Development of the Crisis

Greece and Italy had never achieved the 60 per cent debt to GDP ratio required of membership for the euro – both countries had persistently run ratios over 90 per cent since the 1990s. Other PIIG countries, Ireland, Spain and Portugal, had managed to reduce their debt ratios just below the 60 per cent mark by 2007. At this time, the spread of interest rates across euro area sovereign debt was relatively small suggesting that markets did not anticipate the sort of problems that beset some governments post-crisis. Underlying this apparently benign macroeconomic environment, however, was a sharp rise in borrowing from the private sector in Portugal, Spain, Ireland and Greece. In Greece, domestic credit as a proportion of GDP rose from around 32 per cent in 1998 to 84 per cent in 2007; in Ireland the increase was from around 81 per cent to 184 per cent, in Portugal, from 92 per cent to almost 160 per cent and in Spain from almost 81 per cent to 168.5 per cent. One of the reasons given for this domestic credit boom was that banks could borrow euros on international markets. Prior to the euro, these banks would have had to borrow in other currencies and fluctuations in exchange rates would have made the borrowing more risky. Borrowing was made much easier by the low interest rates which existed throughout much of the early years of the noughties.

This borrowing was used to help finance the boom in housing and construction which took place in each of these countries and helped to drive economic growth. When the financial crisis hit, credit dried up, the housing market collapsed and construction was badly hit. The banking systems across Europe were trying to identify the extent of the exposure that they faced, the potential size of the losses they were likely to make and the possibility of having to borrow to overcome short-term financing problems. Any such borrowing would be difficult or expensive because of the credit-crunch. The reliance of countries like Greece, Ireland, Spain and Portugal on the ability to borrow internationally meant that they were hit hard by the credit crunch. As banks in these countries teetered on the brink of collapse, they looked to national governments for support. If governments allowed these banks to fail then the effect on the population as a whole was likely to be significant, not to mention the wider banking sector in Europe. In order to help support their banking systems, governments had to borrow money – few of these governments had the fiscal flexibility to do so without increasing borrowing.

In addition, the extent of the post-crisis recession had started to affect tax revenues and government spending. In countries like Spain and Ireland, tax revenues decreased as a result of the contraction in the construction industry which had been fuelling pre-crisis growth. This in turn increased the size of budget deficits and with GDP shrinking the size of the deficit to GDP ratio grew. As it became clearer that these governments were likely to breach EU fiscal rules, and with the need to support banking systems growing in intensity, the financial markets' confidence fell and interest rate on the sovereign debt of these countries rose. The gap between the interest rates of the PIIG countries and other EU members widened. Europe had always had countries which were economically sound such as Germany, the Netherlands, Sweden and France and a group of weaker countries. Increasingly, Portugal, Spain, Ireland and Greece came to be referred to as 'the periphery'.

The first part of the Greek bailout was negotiation in late April into May of 2010. In November of 2010, Ireland's government also had to seek support and in May 2011, Portugal followed suit. In June 2011, Greece needed a second bailout which was eventually agreed in March 2012. As part of the agreement, private sector creditors had to accept a 'haircut', losing 50 per cent of the value of their investments. The background of an increasing number of countries being dragged into the crisis and the somewhat chaotic response of the EU to the situation, merely increased nervousness on the financial markets and spreads widened further between the periphery and the core EU countries. The accusation that the response to the crisis had been chaotic led to self-fulfilling speculative attacks from financial markets. High risk countries are more likely to default so investors require higher yields to take on the debt of these countries. If countries have to borrow at higher rates of interest, this in itself increases the risk of default.

Why was the response so chaotic? One reason was the clear divide between those countries that had maintained some degree of fiscal restraint and the periphery which had been seen to free ride on their membership of the EU. Another reason was the realization that political pressures on domestic governments were huge. To satisfy domestic political pressure, the stronger countries such as Germany had to show that it was being firm with these profligate countries. The terms of the bailout packages required the implementation of significant cuts in public spending and tax rises. Some element of fiscal consolidation

was vital to show financial markets that some discipline would be exerted and that the countries seeking support were serious in their willingness to abide by the fiscal rules of the euro.

At the same time, governments knew that implementing austerity packages would be extremely unpopular and impose considerable hardship on the people. Politically, this represented suicide; few governments could be confident of re-election under such circumstances. Opposition parties might promise to stand up to the 'bullies' in the EU who were imposing these policies and to abandon austerity. Such a manifesto might be alluring to people suffering cuts in wages, loss of jobs, cuts to public services, cuts in pension and increases in taxes, but the reality was that even if these opposition groups did find themselves in power, the reality of the situation was not going to go away. If countries did become bankrupt the potential damage to people could have been even worse than the effects of the austerity packages.

In June 2012, Spain had to seek help and in March 2013, the banking system in Cyprus was on the verge of collapse. With banks closed for almost two weeks whilst the Cypriot government negotiated with the EU and IMF, the debate over the extent to which depositors in Cypriot banks should have to accept losses was a key focus. An initial bailout deal imposed a tax on all depositors in an attempt to raise €5.8 billion as part of a bailout deal. The tax was rejected and subsequent negotiations led to an agreement where those with deposits of over €100,000 would be taxed to raise the €5.8 billion contribution to the overall €10 billion bailout. It might seem that those who have such large deposits ought to be the ones shouldering the biggest burden and the amount of deposits from so-called Russian oligarchs in the Cypriot banking system was a feature of the news reporting at the time. However, it will not be just wealthy Russians who, for whatever reason, deposited money in Cypriot banks who will be likely to suffer a 'haircut' – businesses will also be hit. One news report noted that a school deposited fees into its bank which took its total deposits to over the €100,000 threshold. The 'haircut' would mean that it would struggle to pay its bills including wages. Cypriot banks would also be subject to considerable restructuring and whilst banks re-opened there were significant limitations on the amount of money that could be withdrawn and capital controls on money leaving Cyprus were also imposed. At the time of writing it is not clear what the outcome of the Cyprus crisis will be but whilst it is taking place, analysts have begun to look at Slovenia as being the next possible crisis.

The European Financial Stability Fund (EFSF) and the European Stability Mechanism (ESM) The discussions over the Greek bailout and the expectation that more would be needed led European finance ministers to establish the European Financial Stability Fund (EFSF) in May 2010 to provide support for countries that faced default. The EFSF raised funds by issuing bonds and other financial instruments through capital markets. The money raised is lent to countries seeking assistance on the understanding that reforms are put in place. The sum of €750 billion was initially identified as needing to be raised. The initial bailout of Greece accounted for €110 billion of these funds; Ireland's €85 billion and Portugal around €80 billion. The establishment of the EFSF may have resolved some of the more immediate problems that existed in 2010 but merely led to the markets turning their attention to other states deemed to be in danger of default, notably Spain and Italy.

The situation in Italy was of significance because it accounted for a much larger proportion of total EU GDP than Greece. If Italy were to default the consequences for the euro area and the global economy would be dire. This is because Italy is the third largest economy in Europe but has a public debt reported at €1.9 trillion (about five times that of Greece). Between 2011 and 2014, Italy had obligations to redeem around €656 billion in debt. With interest rates on Italian debt at around 6.75 per cent, the cost of servicing this size of debt in a period of economic slump places huge pressure on the government. If interest rates rose by a further 1 per cent, it has been estimated that it would cost Italy a further €38 billion over three years, €78 billion over five years and €313 billion over the life of the debt it has. It might be forced to borrow even more to help service its debt.

In October 2012, the inauguration of the ESM meant that it assumed the tasks of the European Financial Stability Facility (EFSF) and the European Financial Stabilization Mechanism (EFSM). Although the Treaty was signed by the then 17 euro area countries, the ESM will also be open to non-euro area EU countries for ad hoc participation in financial assistance operations. Based in Luxembourg, the ESM is meant to provide a 'permanent crisis resolution mechanism'. The ESM is classed as an intergovernmental organization with a subscribed capital of €700 billion. The 18 members of the euro area are the shareholders and its lending capacity is around €500 billion. Like the EFSF, the ESM will raise funds on the capital markets

and will work with the IMF in dealing with member states who request assistance. The ESM is expected to be part of the solution to member states' problems and will operate in conjunction with the fiscal compact in addressing the fiscal and structural problems facing member states. The ESM will only provide support if the member state seeking help agrees to implement fiscal adjustment and structural reform. Structural reform refers to changes in the labour and capital markets which are designed to help improve the efficiency of the economy. One of the elements of ESM support is a consideration of the situation of the member country and the overall stability of the euro area. If it is considered that the financial stability of the euro area is under threat then the ESM can provide support.

Bailout funds have been used, in part, to help support the banking sectors in each country and help banks recapitalize. Domestic banks, however, typically hold relatively high levels of sovereign debt, if the country is in danger of default then this further weakens the banks' position. Because these countries are part of the single currency and as we have seen, wholesale financial markets are more integrated, the risk of any single country defaulting leading to contagion in other European countries has been a key element of the sovereign debt crisis.

AUSTERITY POLICIES – TOO FAR TOO QUICKLY?

A feature of the sovereign debt crisis has been the adoption by a number of countries of austerity programmes – significant cuts in public spending, tax rises and structural reforms. In principle, it is easy to say that if a country is borrowing too much and has debt problems that it must cut its spending and raise more revenue. The practical implications of this policy are more complex. Not only are the people of the country where austerity programmes are implemented likely to be severely affected, if such policies are adopted during a period of weak economic activity the effects are likely to be that national income will decline and unemployment rise. A deep and lasting recession in Greece, for example, means that the prospects for any growth-led recovery looked bleak; if a country like Greece is experiencing GDP growth rates of –7 per cent then the chances of generating tax revenue to pay off debt and invest in improving the economy are slim. Firms are more likely to fail in such an environment and so corporate borrowing comes under pressure as the risk of default is higher. If corporate bond rates rise then firms will not be able to afford to borrow which further hampers the productive capacity of the economy. The periphery countries find themselves in a very difficult situation.

The UK is also experiencing similar problems. The Coalition Government pledged to cut the deficit and borrowing when it came into office in 2010. Cuts in public spending and tax increases have led to weak economic growth and the economy going back into recession. With the economy performing so badly, the government has found it difficult to see short-term improvements in public finances and re-emphasized the need to stick with its policies for the long-term good of the country. There are many economists who agree that sorting out the UK's debt and deficit problem is important but there are also those that argue that the extent of the austerity measures in a time of weak global economic growth is too much too quickly and have argued that policies focusing on growth are necessary.

Structural and Cyclical Deficits

The focus on public sector deficits raises questions about the difference between cyclical and structural deficits. As we have seen, a cyclical deficit occurs when government spending and income is disrupted by the 'normal' economic cycle. In times of strong economic growth government revenue from taxes will rise and spending on welfare and benefits will fall and so public finances will move into surplus (or the deficit shrinks appreciably). In times of economic slowdown the opposite occurs and the size of the budget deficit will rise (or the surplus shrinks). A structural deficit refers to a situation where the deficit is not dependent on movements in the economic cycle but indicate that a government is 'living beyond its means' – spending what it has not got.

Economists are divided in their views about the nature and importance of structural deficits. What follows is a summary of the arguments on both sides.

Argument 1: Policymakers Need to Eradicate Structural Deficits The existence of a structural deficit implies that the public finances will be even worse when entering a recession, necessitating increasing levels of borrowing which is unsustainable in the long term. Cyclical effects will merely serve to make the deficit worse. This was seen in examples of European government deficits in 2010 – some of the largest deficits in peacetime history and created in part by governments committing to spending too much in the ‘good times’.

As a result of the size of these deficits, the risk of default by governments on their debt is greater; increasingly they will find it difficult to service their debt and the cost of so doing will rise. This also creates uncertainty in financial markets and threatens the survival of the euro. As a result fiscal consolidation is essential to reduce long-term interest rates and currency instability and help to promote economic growth. This fiscal consolidation should primarily be in the form of cuts in public spending rather than increases in taxes which may damage employment and investment.

Argument 2: The Idea of a Structural Deficit is a Myth The idea that government deficits are structural is unhelpful. The assumption is that governments have to borrow more because of the gap between income and expenditure but how certain can we be that the only reason governments are borrowing money is simply due to the changes to public finances wrought by the recession? The whole notion of a structural deficit assumes that it is the amount governments borrow when the economy is operating at its trend level. This implies that to measure it we need to know how far the economy is operating below trend – the output gap.

The problem is that there is considerable disagreement on the size of the output gap. It is accepted that the recession will have destroyed some potential output but how much is open to some interpretation. There have been a number of studies attempting to quantify the impact of economic downturns on the output gap and the outcomes vary significantly. In the UK the Institute for Fiscal Studies (IFS) has estimated that the output loss of the economic downturn from 2008 could be as much as 7.5 per cent whereas the Treasury estimates 5 per cent and other estimates put the gap as low as 2 per cent.

The size of the output gap is important because it has a direct effect on the cyclical component of the deficit – the larger the output gap the larger will be the cyclical component and the smaller the structural. This in turn affects any estimates of the size of borrowing when the economy does return to trend – in other words, the size of the structural deficit. Any calculations on the deficit would also be subject to assumptions about the sensitivity of taxes and spending to changes in GDP. How does tax revenue rise in relation to changes in GDP? This will depend in part on assumptions about the number of people who are able to find work as the economy expands but also on the extent to which potential output has been destroyed. How do changes in government spending vary in relation to changes in GDP? How many people will get off benefits and what effect will short-term fiscal stimulus measures have on spending?

As an increasing number of assumptions are made, calculations of the size of the structural deficit could be very different; which one should be used as the basis for policy decisions? In the light of this analysis is it useful to think of the idea of a structural deficit at all?

SELF TEST Explain why inaccurate information might affect forecasts of economic growth and thus the size of government budget deficits.

Fiscal Consolidation

The extent to which budget deficits should be a cause for concern is a persistent macroeconomic debate. Austerity measures imply that governments should consolidate fiscal policy to reduce deficits and ‘balance the books’. In the UK, for example, the Chancellor of the Exchequer of the Coalition government has regularly referred to ‘balancing the budget’. Whether a government should seek to balance budgets is a source of debate amongst economists. Our study of financial markets showed how budget deficits affect saving, investment and interest rates. But how big a problem are budget deficits?

Argument 1: The Government Should Balance its Budget When a government fails to balance its budget, it has to borrow money by issuing bonds in order to make up the shortfall. The most direct effect

of high and rising government debt is to place a burden on future generations of taxpayers. When these debts and accumulated interest come due, future taxpayers will face a difficult choice. They can pay higher taxes, enjoy less government spending, or both, in order to make resources available to pay off the debt and accumulated interest. Or they can delay the day of reckoning and put the government into even deeper debt by borrowing once again to pay off the old debt and interest. In essence, when the government runs a budget deficit and issues government debt, it allows current taxpayers to pass the bill for some of their government spending on to future taxpayers. Inheriting such a large debt may lower the living standard of future generations.

In addition to this direct effect, budget deficits also have various macroeconomic effects. Because budget deficits represent *negative* public saving, they lower national saving (the sum of private and public saving). Reduced national saving causes real interest rates to rise and investment to fall. Reduced investment leads over time to a smaller stock of capital. A lower capital stock reduces labour productivity, real wages, and the economy's production of goods and services. Thus, when the government increases its debt, future generations are born into an economy with lower incomes as well as higher taxes.

There are, nevertheless, situations in which running a budget deficit is justifiable. Throughout history, the most common cause of increased government debt is war. When a military conflict raises government spending temporarily, it is reasonable to finance this extra spending by borrowing. Otherwise, taxes during wartime would have to rise precipitously. Such high tax rates would greatly distort the incentives faced by those who are taxed, leading to large deadweight losses. In addition, such high tax rates would be unfair to current generations of taxpayers, who already have to make the sacrifice of fighting the war.

Similarly, it is reasonable to allow a budget deficit during a temporary downturn in economic activity. When the economy goes into a recession, tax revenue falls automatically, because income tax and payroll taxes are levied on measures of income, and transfer payments such as unemployment benefit increase. People also spend less so that government income from indirect taxes also falls. If the government tried to balance its budget during a recession, it would have to raise taxes or cut spending at a time of high unemployment. Such a policy would tend to depress aggregate demand at precisely the time it needed to be stimulated and, therefore, would tend to increase the magnitude of economic fluctuations. When the economy goes into recovery, however, the opposite is true: tax receipts rise as the level of economic activity rises and transfer payments tend to fall. The government should therefore be able to run a budget surplus and use the money to pay off the debt incurred by the budget deficit it ran during the recession.

Wars aside, therefore, over the course of the business cycle, there is no excuse for not balancing the budget. If the government runs a deficit when the economy is in a recession, it should run a comparable surplus when the economy recovers, so that on average the budget balances. Compared to the alternative of on-going budget deficits, a balanced budget – or, at least, a budget that is balanced over the business cycle – means greater national saving, investment and economic growth. It means that future university graduates will enter a more prosperous economy.

Argument 2: The Government Should Not Balance its Budget The problem of government debt is often exaggerated. Although government debt does represent a tax burden on younger generations, it is often not large compared with the average person's lifetime income. Often, the case for balancing the government budget is made by confusing the economics of a single person or household with that of a whole economy. Most of us would want to leave some kind of bequest to friends or relatives or a favourite charity when we die – or at least not leave behind large debts. But economies, unlike people, do not have finite lives – in some sense, they live forever, so there is never any reason to clear the debt completely.

Critics of budget deficits sometimes assert that the government debt cannot continue to rise forever, but in fact it can. Just as a bank evaluating a loan application would compare a person's debts to his income, we should judge the burden of the government debt relative to the size of the nation's income. Population growth and technological progress cause the total income of the economy to grow over time. As a result, the nation's ability to pay the interest on the government debt grows over time as well. Really, we should not be looking at the total amount of debt but at the ratio of debt to income. As long as this is not increasing, then the level of debt is sustainable. In other words, as long as the level of government debt grows more slowly than the nation's income, there is nothing to prevent government debt from growing forever. Some numbers can put this into perspective. Suppose the output of the economy grows on average about 3 per cent per year. If the inflation rate averages around 2 per cent per year, then nominal

income grows at a rate of 5 per cent per year. Government debt, therefore, can rise by 5 per cent per year without increasing the ratio of debt to income.

Moreover, it is misleading to view the effects of budget deficits in isolation. The budget deficit is just one piece of a large picture of how the government chooses to raise and spend money. In making these decisions over fiscal policy, policymakers affect different generations of taxpayers in many ways. The government's budget deficit or surplus should be considered together with these other policies. For example, suppose the government reduces the budget deficit by cutting spending on public investments, such as education. Does this policy make young generations better off? The government debt will be smaller when they enter the labour force, which means a smaller tax burden. Yet if they are less well educated than they could be, their productivity and incomes will be lower. Many estimates of the return to schooling (the increase in a worker's wage that results from an additional year in school) find that it is quite large. Reducing the budget deficit rather than funding more education spending could, all things considered, make future generations worse off. A distinction has to be made between borrowing to finance investment, which helps boost the future productive capacity of the economy and borrowing to finance current government expenditure (on things such as wages for public sector workers).

The qualification that the government budget on current expenditure will be balanced on average over the business cycle allows for the effect of automatic stabilizers, such as the increase in welfare expenditure and reduction in tax revenue that automatically occur in a recession (the opposite in a boom) and so helps flatten out economic fluctuations. Allowing a budget deficit on investment expenditure is sensible because, although it leads to rising public debt, it also leads to further growth opportunities through spending on education, roads and so on. Just asserting that the government should balance its budget, irrespective of the economic cycle and irrespective of what kind of expenditures it is making, is overly simplistic.

SELF TEST Why might we wish to distinguish between government current expenditure and government investment expenditure and take account of the economic cycle when judging whether the government should balance its budget?

Austerity or Growth?

The arguments outlined above crystallize the debate over the importance of austerity policies. Whilst there might be general agreement that countries do need to get more control over debt and their deficits, the current policies being adopted are, some argue, counter-productive.

In the UK, the Chancellor has had to announce further cuts in public spending and observed that austerity would have to last much longer than he originally thought. Opposition politicians have argued that the government went too far too quickly with its austerity measures and that the continued weakness in the economy was testament to the fact that austerity was not working to reduce the deficit and debt. Why the difference?

One reason is the assumptions that underpin the models that are used in calculating the effects of austerity. In the UK, the independent Office for Budget Responsibility (OBR) based its calculations on the assumption of a fiscal multiplier of 0.5. In other words, every £1 of cuts in public spending would reduce economic output in the economy by 50p. The IMF have estimated the size of these fiscal multipliers to be much higher, at somewhere between 0.9–1.7. At the upper end of the estimates of the fiscal multiplier the effect on the economy is considerable. Estimates have put the effect of the government's cuts at around 8 per cent of GDP over five years if the fiscal multiplier is assumed to be 1.3, about the middle of the IMF's range. The government of the UK would argue that even if the size of the fiscal multiplier was at this level, some of the effects would be offset by the level of quantitative easing by the Bank of England.

The Argument for Austerity Countries must take steps to get public finances under control. If attempts are made to increase borrowing to spend their way out of recession the effect will be increased inflation, higher borrowing costs across the whole of the EU and further uncertainty which threatens the whole of the euro area. There is also a moral dimension to the necessity for austerity – it is not fair to expect the

taxpayers of countries which have maintained a sound fiscal stance and abided by the rules of the euro area to pay for the lax behaviour of governments in other countries.

In the short to medium term the effects of austerity are considerable but the long-term benefits of sounder finances, improved structural reforms and a stronger banking sector will mean that all countries in Europe will be more competitive and in a position to benefit from stronger and more sustained economic growth in the future.

The Argument for Growth The main cause of the problems facing Europe stem from the financial system not from government debt. Austerity at a time of economic slowdown will affect economies considerably and will be further exacerbated by the lack of credit in economies caused by banks seeking to recapitalize and rebuild their balance sheets. Without an adequate supply of credit, consumer spending and business investment will decline and lead to a prolonged and damaging period of economic depression. If output declines, tax revenues will fall and government spending on benefits will rise and far from alleviating the debt crisis will merely mean governments are forced to borrow more. The only way to solve the debt crisis is to get economies growing again. Governments need to adopt Keynesian style stimulus spending; stronger economic growth will lead to increased tax revenues and if unemployment falls, government spending on welfare support drops reducing the need to borrow.

A Resolution?

At the time of writing the outcome of the debate over austerity and growth is far from resolved and indeed it is probably unfair to suggest that the argument is a simple split between those pushing to maintain austerity at all costs and those who want to abandon austerity and go for growth. The EU Commission issued a statement in the Spring of 2013 refuting suggestions that austerity policies lacked flexibility and that the necessity for growth was being ignored.

The Commission argues that the extent of the spreads on sovereign debt did indicate that there were legitimate concerns over the threat of default by some countries but acknowledge that there were also some speculative reasons for the extent of the spreads. It also points to the policy adopted by the ECB of outright monetary transactions (OMT) where the ECB buys the bonds of Europe's indebted countries to help push up the price and lower yields. This will not only help reduce the cost of servicing the debt but also signals to the markets that the euro will be protected. To make such a policy credible, structural reforms and fiscal consolidation is necessary. Structural reforms, such as improving the flexibility of labour and product markets will help to reduce wage and price rigidities, will take time to yield benefits but are essential if Europe is to be competitive in a global economy. The lowering of yields and spreads on sovereign debt in the latter part of 2012 and into 2013 was testament, it argued, that the policy was starting to have some impact.

IN THE NEWS



Sovereign Debt

Three years after the height of the sovereign debt crisis, are the sacrifices made by the countries who asked for a bailout and the EU showing any signs of bearing fruit?

The PIIGS – Three Years On

The first quarter of 2013 saw some optimism returning to financial markets around the world. The stock market in the US rose as did the London Stock Exchange. Bullish sentiment on stock prices was driven in part by more optimistic economic news coming out of the US and by the feeling that the worst of the sovereign debt crisis had been weathered. The banking crisis in Cyprus did set back sentiment but the relatively speedy agreement to organize a bailout seemed to calm nerves once again. What of the rest of the euro area? Is the cause for optimism backed up by what is happening in sovereign debt markets?

In the latter part of March 2013, Italy sold €3.9 billion in five-year debt but the yield rose from 3.59 per cent to 3.65 per cent, the highest since October 2012 and analysts noted that the demand for the debt was 'soft'. Italy also sold €3.0 billion of ten-year debt and the yield here fell from

One of the major fears of the problems which originated in Greece was contagion – the potential for the euro crisis to spread to the weaker economies of the euro area and cause the collapse of the euro.



4.83 per cent to 4.66 per cent. In the same month, Spain sold €5.83 billion of three- and six-month treasury bills amidst stronger demand and a lower yield (0.79 per cent compared to 0.85 per cent).

At the same time, Greek banks announced losses, after a €100 billion write-down in sovereign debt. The National Bank reported losses of €2.14 billion for the year in March 2013, Alpha Bank reported losses of €1.09 billion. Both banks are part of the recapitalization process that is a term of the continued funding arrangements from the EU. Both banks need to ensure that at least 10 per cent of any new capital they secure comes from private investors. Ireland, meanwhile, managed to sell debt on the markets for the first time since 2010 and the Irish Taoiseach, Enda Kenny, noted that the country had made considerable progress in its plans to leave the bailout programme it agreed to in 2010. Part of the Irish government's plans include ensuring that promises made by the EU to help Ireland develop initiatives such as guaranteeing young people the right to receive training are acted upon. The Taoiseach said that '... it's absolutely critical for governments to be able to bring people with them and to explain both the nature of the problem, the scale of that problem, the plan and the strategy to deal with it'.

The ratio of public debt to GDP in Portugal rose between the second and third quarters of 2012 and reached 120.3 per cent, up from 117.4 per cent. Yields on Portuguese

debt are falling, however; ten-year government bonds attracted a yield of 12.5 per cent in mid-April 2012 but a year later yields were hovering nearer the 6 per cent mark. Ratings agencies, however, confirmed that they retained a negative outlook on Portuguese debt with Moody's rating it at Ba3 (the equivalent of BB- or non-investment grade, speculative).

Questions

- 1 If optimism leads to increases in average stock prices, does this mean that there is also optimism that the sovereign debt crisis is over the worst? Explain.**
- 2 What does the movement in the yield of Italian five- and ten-year debt tell you about the price and the view of the markets of Italian debt?**
- 3 Why do you think that the Irish Taoiseach placed an emphasis on it being '... absolutely critical for governments to be able to bring people with them and to explain both the nature of the problem, the scale of that problem, the plan and the strategy to deal with it' in managing an orderly exit from an EU bailout?**
- 4 Can any conclusions be drawn on the movement of the size of the debt in Portugal, its yields and the ratings on the state of the Portuguese economy?**
- 5 To what extent do you think that the terms of the bailouts of countries like Greece, Ireland and Portugal have been justified?**

SUMMARY

- Deregulation of financial markets encouraged new ways of lending and made it easier for lenders to access funds and the relatively benign economic climate of the early 2000s encouraged banks to become more risk seeking.
- Banks around the world increased lending rapidly and this helped to fuel a rise in property prices. New financial products meant that risk was perceived to be lower but many involved did not have a complete understanding of the complexity of the products or the models on which they were based.
- When central banks began to start raising interest rates to control accelerating inflation, the number of defaults on sub-prime mortgages began to increase and as the web on interconnectivity between financial institutions began to unravel, credit became tight.
- The banking crisis led to a sharp reduction in interbank lending as credit dried up which fed through the housing market causing a collapse in house prices and quickly fed through to the real economy leading to rising unemployment, shrinking economic activity.
- When the crisis hit, central banks responded by reducing interest rates and developing new techniques to complement fiscal stimuli such as quantitative easing.
- Both central banks and regulators have come in for criticism about their role in the crisis and their response after and reforms to banking systems and regulation continue to be discussed.
- The sovereign debt crisis developed as a result of the financial crisis. Countries experiencing economic slow-down were forced to borrow more and the extent of their debt created nervousness of the prospect of default.
- The spreads on sovereign debt widened into 2010 and 2011 and a number of countries had to seek financial assistance from the EU and IMF. The ESM represents a permanent mechanism to help support countries in financial difficulties.
- Part of the conditions for receiving assistance has been the implementation of austerity programmes which have further impacted on economic growth across the EU.
- A debate between the benefits of austerity and the need to promote growth has developed which is on-going.

QUESTIONS FOR REVIEW

- 1 What is meant by 'deregulation of the financial markets'?
- 2 How do 'bubbles' in asset prices occur?
- 3 What were the main causes of the rise in asset prices in the run up to the financial crisis?
- 4 What were the main causes of the financial crisis?
- 5 Why did the financial crisis have such an impact on global economic activity?
- 6 How did central banks respond to the financial crisis?
- 7 Why did governments such as those in Greece, Spain, Ireland, Portugal and Italy experience such a rise in deficits and debt in the wake of the financial crisis?
- 8 What is meant by the term 'contagion' in the context of the sovereign debt crisis?
- 9 What are the main arguments for the continuation of austerity programmes in countries which have experienced sovereign debt problems?
- 10 What are the main arguments against austerity programmes as a solution to the sovereign debt crisis?

PROBLEMS AND APPLICATIONS

- 1 Explain how deregulation of the financial system may have affected the following:
 - a. a newly married couple aged 24 who have just secured their first jobs and want to buy a house
 - b. joint homeowners with a house valued at €350,000 with a mortgage of €200,000 who want to celebrate their 25th wedding anniversary with a world cruise and who are looking to borrow money to do so
 - c. the market for newly constructed properties.
- 2 In the face of a financial crisis such as that created by the dot.com collapse and the 9/11 terrorist attacks in the United States, why would a central bank have

responded by lowering interest rates and flooding the markets with liquidity?

- 3 Eugene Fama, a key figure in the development of the theory of the efficient markets hypothesis, commented in an interview with the *New Yorker* (13 January 2010):

'If a bubble means an extended period during which asset prices depart quite significantly from economic fundamentals that means that somebody must have made a lot of money betting on that, if you could identify it. It's easy to say prices went down, it must have been a bubble, after the fact. I think most bubbles are twenty-twenty hindsight. Now after the fact you always find people who said before the fact that prices are too high. People are always saying that prices are too high. When they turn out to be right, we anoint them. When they turn out to be wrong, we ignore them. They are typically right and wrong about half the time.'

To what extent do you think that the rise in asset prices leading up to the financial crisis was a 'bubble' or that in the aftermath of the crisis reference to a bubble was just twenty-twenty hindsight?

- 4 To what extent do you think that the response by central banks to the financial crisis was 'fragmented'?

- 5 Why have some critics argued that the cause of the sovereign debt crisis lies in the poor design of the euro system and the ability of countries within the euro area to free ride?
- 6 What is the role of the European Stability Mechanism in managing sovereign debt problems of member countries?
- 7 Is there any meaningful difference between a cyclical deficit and a structural deficit? Explain your answer.
- 8 Governments across Europe have instituted austerity measures in an attempt to cut budget deficits which ballooned after the financial crisis. Are they right to do this at a time when the European economic recovery is acknowledged as being fragile?
- 9 Suppose the government cuts taxes and increases spending, raising the budget deficit to 12 per cent of GDP. If nominal GDP is rising 7 per cent per year, are such budget deficits sustainable forever? Explain. If budget deficits of this size are maintained for 20 years, what is likely to happen to your taxes and your children's taxes in the future? Can you do something today to offset this future effect?
- 10 The chapter says that budget deficits reduce the income of future generations, but can boost output and income during a recession. Explain how both of these statements can be true.

GLOSSARY

- ability-to-pay principle** the idea that taxes should be levied on a person according to how well that person can shoulder the burden
- abnormal profit** the profit over and above normal profit
- absolute advantage** absolute advantage exists where a producer can produce a good using fewer factor inputs than another
- absolute poverty** a level of poverty where an individual does not have access to the basics of life – food, clothing and shelter
- accounting profit** total revenue minus total explicit cost
- actual spending, saving or investment** the realized or *ex post* outcome resulting from actions of households and firms
- ad valorem tax** a tax levied as a percentage of the price of a good
- adaptive expectations** a model which states that individuals and organizations base their expectations of inflation in the future on past actual inflation rates
- adverse selection** where a principal knows more about their situation than the agent, leading to the agent preferring not to do business with the principal
- agent** a person who is performing an act for another person, called the principal
- aggregate demand curve** a curve that shows the quantity of goods and services that households, firms and the government want to buy at each price level
- aggregate risk** risk that affects all economic actors at once
- aggregate supply curve** a curve that shows the quantity of goods and services that firms choose to produce and sell at each price level
- allocative efficiency** a resource allocation where the value of the output by sellers matches the value placed on that output by buyers
- amplitude** the difference between peak and trough and trend output
- appreciation** an increase in the value of a currency as measured by the amount of foreign currency it can buy
- arbitrage** a trade which seeks to exploit price differences in order to make profit through the process of buying a good in one market at a low price and selling it in another market at a higher price in order to profit from the price difference
- asymmetric information** where two parties have access to different information
- asymmetric shocks** a situation where changes in aggregate demand and/or supply differ from one country to another
- automatic stabilizers** changes in fiscal policy that stimulate aggregate demand when the economy goes into a recession, without policymakers having to take any deliberate action
- autonomous spending or autonomous expenditure** spending which is not dependent on income/output
- average fixed cost** fixed costs divided by the quantity of output
- average revenue** total revenue divided by the quantity sold
- average tax rate** total taxes paid divided by total income
- average total cost** total cost divided by the quantity of output
- average variable cost** variable costs divided by the quantity of output
- balanced budget** where the total sum of money received by a government in tax revenue and interest is equal to the amount it spends, including on any debt interest owing
- balanced trade** a situation in which exports equal imports
- Bank of England** the central bank of the United Kingdom
- barriers to entry** anything which prevents a firm from entering a market or industry
- barter** the exchange of one good or service for another
- benefits principle** the idea that people should pay taxes based on the benefits they receive from government services
- bond** a certificate of indebtedness
- bounded rationality** the idea that humans make decisions under the constraints of limited, and sometimes unreliable, information
- brain drain** the emigration of many of the most highly educated workers to rich countries
- branding** the means by which a business creates an identity for itself and highlights the way in which it differs from its rivals
- budget constraint** the limit on the consumption bundles that a consumer can afford
- business cycle** the fluctuations in economic growth around trend growth
- capital** the equipment and structures used to produce goods and services
- capital flight** a large and sudden reduction in the demand for assets located in a country
- cartel** a group of firms acting in unison
- catch-up effect** the property whereby countries that start off poor tend to grow more rapidly than countries that start off rich
- central bank** an institution designed to regulate the quantity of money in the economy
- choice set** the set of alternatives available to the consumer
- circular-flow diagram** a visual model of the economy that shows how money and production inputs and outputs flow through markets among households and firms
- classical dichotomy** the theoretical separation of nominal and real variables
- closed economy** an economy that does not interact with other economies in the world
- Coase theorem** the proposition that if private parties can bargain without cost over the allocation of resources, they can solve the problem of externalities on their own
- coincident indicator** an indicator which occurs at the same time as changes in economic activity
- collective bargaining** the process by which unions and firms agree on the terms of employment
- collusion** an agreement among firms in a market about quantities to produce or prices to charge
- commodity money** money that takes the form of a commodity with intrinsic value

- common currency area** a geographical area throughout which a single currency circulates as the medium of exchange
- common resources** goods that are rival but not excludable
- comovement** the movement of pairs of variables over time
- comparative advantage** the comparison among producers of a good according to their opportunity cost. A producer is said to have a comparative advantage in the production of a good if the opportunity cost is lower than that of another producer
- compensating differential** a difference in wages that arises to offset the non-monetary characteristics of different jobs
- competitive advantage** the advantages firms can gain over another which have the characteristics of being both distinctive and defensible
- competitive market** a market in which there are many buyers and sellers so that each has a negligible impact on the market price
- complements** two goods for which an increase in the price of one leads to a decrease in the demand for the other
- compounding** the accumulation of a sum of money in, say, a bank account, where the interest earned remains in the account to earn additional interest in the future
- concentration ratio** the proportion of total market share accounted for by a particular number of firms
- constant returns to scale** the property whereby long-run average total cost stays the same as the quantity of output changes
- consumer prices index** a measure of the overall prices of the goods and services bought by a typical consumer
- consumer surplus** a buyer's willingness to pay minus the amount the buyer actually pays
- consumption** spending by households on goods and services, with the exception of purchases of new housing
- contraction** when real output is lower than the previous time period
- copyright** the right of an individual or organization to own things they create in the same way as a physical object to prevent others from copying or reproducing the creation
- cost** the value of everything a seller must give up to produce a good
- cost-benefit analysis** a study that compares the costs and benefits to society of providing a public good
- countercyclical** a variable that is below trend when GDP is above trend
- cream-skimming** a situation where a firm identifies parts of a market that are high in value added and seeks to exploit those markets
- credit default swap (CDS)** a means by which a bondholder can insure against the risk of default
- cronyism** a situation where the allocation of resources in the market is determined in part by political decision making and favours rather than by economic forces
- cross-price elasticity of demand** a measure of how much the quantity demanded of one good responds to a change in the price of another good, computed as the percentage change in quantity demanded of the first good divided by the percentage change in the price of the second good
- crowding out** a decrease in investment that results from government borrowing
- crowding-out effect** the offset in aggregate demand that results when expansionary fiscal policy raises the interest rate and thereby reduces investment spending
- currency** the paper banknotes and coins in the hands of the public
- cyclical deficit** a situation when government spending and income is disrupted by the deviations in the 'normal' economic cycle
- cyclical unemployment** the deviation of unemployment from its natural rate
- deadweight loss** the fall in total surplus that results from a market distortion, such as a tax
- deflationary gap or output gap** the difference between full employment output and expenditure when expenditure is less than full employment output
- demand curve** a graph of the relationship between the price of a good and the quantity demanded
- demand deposits** balances in bank accounts that depositors can access on demand by using a debit card or writing a cheque
- demand schedule** a table that shows the relationship between the price of a good and the quantity demanded
- de-merit goods** goods that are over-consumed if left to the market mechanism and which generate both private and social costs which are not taken into account by the decision maker
- depreciation** a decrease in the value of a currency as measured by the amount of foreign currency it can buy
- depression** a severe recession
- derived demand** a situation where demand is determined by the supply in another market
- deterministic trends** trends that are constant, positive or negative independent of time for the series being analysed
- diminishing marginal product** the property whereby the marginal product of an input declines as the quantity of the input increases
- diminishing marginal utility** refers to the tendency for the additional satisfaction from consuming extra units of a good to fall
- direct tax** a tax levied on income and wealth
- discount rate** the interest rate at which the Federal Reserve lends on a short-term basis to the US banking sector
- discrimination** the offering of different opportunities to similar individuals who differ only by race, ethnic group, sex, age or other personal characteristics
- diseconomies of scale** the property whereby long-run average total cost rises as the quantity of output increases
- diversification** the reduction of risk achieved by replacing a single risk with a large number of smaller unrelated risks
- dominant strategy** a strategy that is best for a player in a game regardless of the strategies chosen by the other players
- double coincidence of wants** a situation in exchange where two people each have a good or service that the other wants and can thus enter into an exchange
- economic activity** how much buying and selling goes on in the economy over a period of time
- economic bubble** when prices of assets and securities rise way above their true or fundamental value

- economic growth** the increase in the amount of goods and services in an economy over a period of time
- economic profit** total revenue minus total cost, including both explicit and implicit costs
- economic rent** the amount a factor of production earns over and above its transfer earnings
- economics** the study of how society manages its scarce resources
- economies of scale** the property whereby long-run average total cost falls as the quantity of output increases
- efficiency** the property of a resource allocation of maximizing the total surplus received by all members of society
- efficiency wages** above-equilibrium wages paid by firms in order to increase worker productivity
- efficient markets hypothesis** the theory that asset prices reflect all publicly available information about the value of an asset
- efficient scale** the quantity of output that minimizes average total cost
- elasticity** a measure of the responsiveness of quantity demanded or quantity supplied to one of its determinants
- endogenous variable** a variable whose value is determined within the model
- endowment effect** where the value placed on something owned is greater than on an identical item not owned
- Engel curve** a line showing the relationship between demand and levels of income
- entry limit pricing** a situation where a firm will keep prices lower than they could be in order to deter new entrants
- equilibrium or market price** the price where the quantity demanded is the same as the quantity supplied
- equilibrium quantity** the quantity bought and sold at the equilibrium price
- equity** the property of distributing economic prosperity fairly among the members of society
- European Central Bank (ECB)** the overall central bank of the 18 countries comprising the European Monetary Union
- European Economic and Monetary Union** the European currency union that has adopted the euro as its common currency
- European Union** a family of democratic European countries committed to working together for peace and prosperity
- Eurosystem** the system made up of the ECB plus the national central banks of each of the 18 countries comprising the European Monetary Union
- excludable** the property of a good whereby a person can be prevented from using it when they do not pay for it
- exogenous variable** a variable whose value is determined outside the model
- expected utility theory** the idea that preferences can and will be ranked by buyers
- explicit costs** input costs that require an outlay of money by the firm
- exports** goods produced domestically and sold abroad leading to an inflow of funds into a country
- externality** the cost or benefit of one person's decision on the well-being of a bystander (a third party) which the decision maker does not take into account in making the decision
- fiat money** money without intrinsic value that is used as money because of government decree
- finance** the field of economics that studies how people make decisions regarding the allocation of resources over time and the handling of risk
- financial intermediaries** financial institutions through which savers can indirectly provide funds to borrowers
- financial markets** financial institutions through which savers can directly provide funds to borrowers
- financial system** the group of institutions in the economy that help to match one person's saving with another person's investment
- fiscal federalism** a fiscal system for a group of countries involving a common fiscal budget and a system of taxes and fiscal transfers across countries
- Fisher effect** the one-for-one adjustment of the nominal interest rate to the inflation rate
- fixed costs** costs that are not determined by the quantity of output produced
- foreign direct investment** capital investment that is owned and operated by a foreign entity
- foreign portfolio investment** investment that is financed with foreign money but operated by domestic residents
- fractional-reserve banking** a banking system in which banks hold only a fraction of deposits as reserves
- free rider** a person who receives the benefit of a good but avoids paying for it
- frictional unemployment** unemployment that results because it takes time for workers to search for the jobs that best suit their tastes and skills
- full employment** a point where those people who want to work at the going market wage level are able to find a job
- fundamental analysis** the study of a company's accounting statements and future prospects to determine its value
- future value** the amount of money in the future that an amount of money today will yield, given prevailing interest rates
- game theory** the study of how people behave in strategic situations
- GDP at constant prices** gross domestic product calculated using prices that existed at a particular base year which takes into account changes in inflation over time
- GDP at current or market prices** gross domestic product calculated by multiplying the output of goods and services by the price of those goods and services in the reporting year
- GDP deflator** a measure of the price level calculated as the ratio of nominal GDP to real GDP times 100
- GDP per capita** gross domestic product divided by the population of a country to give a measure of national income per head
- general equilibrium** the notion that the decisions and choices of economic agents are coordinated across markets
- Giffen good** a good for which an increase in the price raises the quantity demanded
- Gini coefficient** a measure of the degree of inequality of income in a country
- government failure** a situation where political power and incentives distort decision making so that decisions are made which conflict with economic efficiency

government purchases spending on goods and services by local, state and national governments

gross domestic product (GDP) the market value of all final goods and services produced within a country in a given period of time

gross domestic product per capita (head) the market value of all goods and services produced within a country in a given period of time divided by the population of a country to give a per capita figure

gross investment the total spending on capital stock per period of time in an economy

heuristics short cuts or rules of thumb that people use in decision making

horizontal equity the idea that taxpayers with similar abilities to pay taxes should pay the same amount

human capital the accumulation of investments in people, such as education and on-the-job training

hysteresis the lagging effects of past economic events on future ones

idiosyncratic risk risk that affects only a single economic actor

imperfect competition exists where firms are able to differentiate their product in some way and so can have some influence over price

implicit costs input costs that do not require an outlay of money by the firm

import quota a limit on the quantity of a good that can be produced abroad and sold domestically

imports goods produced abroad and purchased for use in the domestic economy leading to an outflow of funds from a country

income effect the change in consumption that results when a price change moves the consumer to a higher or lower indifference curve

income elasticity of demand a measure of how much the quantity demanded of a good responds to a change in consumers' income, computed as the percentage change in quantity demanded divided by the percentage change in income

indexation the automatic correction of a money amount for the effects of inflation by law or contract

indifference curve a curve that shows consumption bundles that give the consumer the same level of satisfaction

indirect tax a tax levied on the sale of goods and services

inferior good a good for which, *ceteris paribus*, an increase in income leads to a decrease in demand (and vice versa)

inflation an increase in the overall level of prices in the economy

inflation rate the percentage change in the price index from the preceding period

inflation tax the revenue the government raises by creating money

inflationary gap the difference between full employment output and actual expenditure when actual expenditure is greater than full employment output

informationally efficient reflecting all available information in a rational way

in-kind transfers transfers to the poor given in the form of goods and services rather than cash

internalizing an externality altering incentives so that people take account of the external effects of their actions

intertemporal choice where decisions made today can affect choices facing individuals in the future

interventionist supply-side policies policies focused on improving the working of markets through investing in infrastructure, education and research and development

investment spending on capital equipment, inventories and structures, including household purchases of new housing

investment fund an institution that sells shares to the public and uses the proceeds to buy a portfolio of stocks and bonds

isocost line a line showing the different combination of factor inputs which can be purchased with a given budget

job search the process by which workers find appropriate jobs given their tastes and skills

labour the human effort both mental and physical that goes in to production

labour force the total number of workers, including both the employed and the unemployed

labour force participation rate (or economic activity rate) the percentage of the adult population that is in the labour force

Laffer curve the relationship between tax rates and tax revenue

lagging indicator an indicator which occurs after changes in economic activity have occurred

land all the natural resources of the earth

law of demand the claim that, other things equal (*ceteris paribus*) the quantity demanded of a good falls when the price of the good rises

law of supply the claim that, *ceteris paribus*, the quantity supplied of a good rises when the price of a good rises

law of supply and demand the claim that the price of any good adjusts to bring the quantity supplied and the quantity demanded for that good into balance

leading indicator an indicator which can be used to foretell future changes in economic activity

liberalism the political philosophy according to which the government should choose policies deemed to be just, as evaluated by an impartial observer behind a 'veil of ignorance'

libertarianism the political philosophy according to which the government should punish crimes and enforce voluntary agreements but not redistribute income

life cycle the regular pattern of income variation over a person's life

liquidity the ease with which an asset can be converted into the economy's medium of exchange

logrolling the agreement between politicians to exchange support on an issue

Lorenz curve the relationship between the cumulative percentage of households and the cumulative percentage of income

lump-sum tax a tax that is the same amount for every person

macroeconomics the study of economy-wide phenomena, including inflation, unemployment and economic growth

marginal abatement cost the cost expressed in terms of the last unit of pollution not emitted (abated)

marginal changes small incremental adjustments to a plan of action

- marginal cost** the increase in total cost that arises from an extra unit of production
- marginal product** the increase in output that arises from an additional unit of input
- marginal product of labour** the increase in the amount of output from an additional unit of labour
- marginal propensity to consume** the fraction of extra income that a household consumes rather than saves
- marginal propensity to save** the fraction of extra income that a household saves rather than consumes
- marginal rate of substitution** the rate at which a consumer is willing to trade one good for another
- marginal rate of technical substitution** the rate at which one factor input can be substituted for another at a given level of output
- marginal revenue** the change in total revenue from an additional unit sold
- marginal revenue product** the extra revenue a firm gets from hiring an additional unit of a factor of production
- marginal tax rate** the extra taxes paid on an additional unit of income
- marginal utility** the addition to total utility as a result of consuming one extra unit of a good
- market** a group of buyers and sellers of a particular good or service
- market economy** an economy that addresses the three key questions of the economic problem through allocating resources through the decentralized decisions of many firms and households as they interact in markets for goods and services
- market failure** a situation where scarce resources are not allocated to their most efficient use
- market for loanable funds** the market in which those who want to save supply funds and those who want to borrow to invest demand funds
- market-orientated supply-side policies** policies designed to free up markets to improve resource allocation through more effective price signals
- market power** the ability of a single economic agent (or small group of agents) to have a substantial influence on market prices
- market segments** the breaking down of customers into groups with similar buying habits or characteristics
- market share** the proportion of total sales in a market accounted for by a particular firm
- maximin criterion** the claim that the government should aim to maximize the well-being of the worst-off person in society
- medium of exchange** an item that buyers give to sellers when they want to purchase goods and services
- menu costs** the costs of changing prices
- merit goods** goods which can be provided by the market but may be under-consumed as a result
- microeconomics** the study of how households and firms make decisions and how they interact in markets
- minimum wage** the lowest price an employer may legally pay to a worker
- model of aggregate demand and aggregate supply** the model that most economists use to explain short-run fluctuations in economic activity around its long-run trend
- monetary neutrality** the proposition that changes in the money supply do not affect real variables
- monetary policy** the set of actions taken by the central bank in order to affect the money supply
- money** the set of assets in an economy that people regularly use to buy goods and services from other people
- money market** the market in which the commercial banks lend money to one another on a short-term basis
- money multiplier** the amount of money the banking system generates with each unit of reserves
- money stock** the quantity of money circulating in the economy
- money supply** the quantity of money available in the economy
- monopolistic competition** a market structure in which many firms sell products that are similar but not identical
- monopoly** a firm that is the sole seller of a product without close substitutes
- monopsony** a market with a single buyer
- moral hazard** the tendency of a person who is imperfectly monitored to engage in dishonest or otherwise undesirable behaviour
- multiplier effect** the additional shifts in aggregate demand that result when expansionary fiscal policy increases income and thereby increases consumer spending
- Nash equilibrium** a situation in which economic actors interacting with one another each choose their best strategy given the strategies that all the other actors have chosen
- national saving (saving)** the total income in the economy that remains after paying for consumption and government purchases
- natural monopoly** a monopoly that arises because a single firm can supply a good or service to an entire market at a smaller cost than could two or more firms
- natural-rate hypothesis** the claim that unemployment eventually returns to its normal, or natural, rate, regardless of the rate of inflation
- natural rate of output** the output level in an economy when all existing factors of production (land, labour, capital and technology resources) are fully utilized and where unemployment is at its natural rate
- natural rate of unemployment** the normal rate of unemployment around which the unemployment rate fluctuates
- natural resources** the inputs into the production of goods and services that are provided by nature, such as land, rivers and mineral deposits
- negative externality** the costs imposed on a third party of a decision
- negative income tax** a tax system that collects revenue from high-income households and gives transfers to low-income households
- net capital outflow** the purchase of foreign assets by domestic residents minus the purchase of domestic assets by foreigners
- net exports** spending on domestically produced goods by foreigners (exports) minus spending on foreign goods by domestic residents (imports)
- net investment** spending on the capital stock taking into account spending on depreciation of the existing capital stock

nominal exchange rate the rate at which a person can trade the currency of one country for the currency of another

nominal interest rate the interest rate as usually reported without a correction for the effects of inflation

nominal variables variables measured in monetary units

non-price competition a situation where two or more firms seek to increase demand and market share by methods other than through changing price

nonstationary data time series data where the mean value can either rise or fall over time

normal good a good for which, *ceteris paribus*, an increase in income leads to an increase in demand (and vice versa)

normal profit the minimum amount required to keep factors of production in their current use

normative statements claims that attempt to prescribe how the world should be

Okun's law a 'law' which is based on observations that in order to keep the unemployment rate steady, real GDP needs to grow at or close to its potential

oligopoly competition amongst the few – a market structure in which only a few sellers offer similar or identical products and dominate the market

open economy an economy that interacts freely with other economies around the world

open-market operations the purchase and sale of non-monetary assets from and to the banking sector by the central bank

opportunity cost whatever must be given up to obtain some item; the value of the benefits foregone (sacrificed)

optimum currency area a group of countries for which it is optimal to adopt a common currency and form a currency union

outright open-market operations the outright sale or purchase of non-monetary assets to or from the banking sector by the central bank without a corresponding agreement to reverse the transaction at a later date

Pareto improvement when an action makes at least one economic agent better off without harming another economic agent

patent the right conferred on the owner to prevent anyone else making or using an invention or manufacturing process without permission

payoff matrix a table showing the possible combination of outcomes (payoffs) depending on the strategy chosen by each player

peak where economic activity reaches a high and real output begins to decline

perfect complements two goods with right-angle indifference curves

perfect price discrimination a situation in which the monopolist knows exactly the willingness to pay of each customer and can charge each customer a different price

perfect substitutes two goods with straight-line indifference curves

permanent income a person's normal income

Phillips curve a curve that shows the short-run trade-off between inflation and unemployment

physical capital the stock of equipment and structures that are used to produce goods and services

Pigovian tax a tax enacted to correct the effects of a negative externality

planned spending, saving or investment the desired or intended actions of households and firms

positional arms races a situation where individuals invest in a series of measures designed to gain them an advantage but which simply offset each other

positional externality a situation which exists when the payoff to one individual is dependent on their relative performance to others

positive externality the benefits to a third party of a decision

positive statements claims that attempt to describe the world as it is

poverty line an absolute level of income set by the government below which a family is deemed to be in poverty. In the UK and Europe this is measured by earnings less than 60 per cent of median income

poverty rate the percentage of the population whose family income falls below an absolute level called the poverty line

predatory or destroyer pricing a situation where firms hold price below average cost for a period to try and force out competitors or prevent new firms from entering the market

present value the amount of money today that would be needed to produce, using prevailing interest rates, a given future amount of money

price ceiling a legal maximum on the price at which a good can be sold

price-consumption curve a line showing the consumer optimum for two goods as the price of one of the goods changes assuming incomes and the price of the good are held constant

price discrimination the business practice of selling the same good at different prices to different customers

price elasticity of demand a measure of how much the quantity demanded of a good responds to a change in the price of that good, computed as the percentage change in quantity demanded divided by the percentage change in price

price elasticity of supply a measure of how much the quantity supplied of a good responds to a change in the price of that good, computed as the percentage change in quantity supplied divided by the percentage change in price

price floor a legal minimum on the price at which a good can be sold

price level an average of prices of all goods and services in an economy as measured by the CPI or the GDP deflator

principal a person for whom another person, called the agent, is performing some act

prisoners' dilemma a particular 'game' between two captured prisoners that illustrates why cooperation is difficult to maintain even when it is mutually beneficial

private goods goods that are both excludable and rival

private saving the income that households have left after paying for taxes and consumption

private sector that part of the economy where business activity is owned, financed and controlled by private individuals

privatization the transfer of publicly owned assets to private sector ownership

procyclical a variable that is above trend when GDP is above trend

- producer prices index** a measure of the change in prices of a basket of goods and services bought by firms
- producer surplus** the amount a seller is paid for a good minus the seller's cost
- production function** the relationship between the quantity of inputs used to make a good and the quantity of output of that good
- production isoquant** a function representing all possible combinations of factor inputs that can be used to produce a given level of output
- production possibilities frontier** a graph that shows the combinations of output that the economy can possibly produce given the available factors of production and available production technology
- productivity** the quantity of goods and services produced from each hour of a worker or factor of production's time
- progressive tax** a tax for which high-income taxpayers pay a larger fraction of their income than do low-income taxpayers
- property rights** the exclusive right of an individual, group or organization to determine how a resource is used
- proportional or flat tax** a tax for which high-income and low-income taxpayers pay the same fraction of income
- prospect theory** a theory that suggests people attach different values to gains and losses and do so in relation to some reference point
- public choice theory** the analysis of governmental behaviour, and the behaviour of individuals who interact with government
- public goods** goods that are neither excludable nor rival
- public interest** making decisions based on a principle where the maximum benefit is gained by the largest number of people at minimum cost
- public saving** the tax revenue that the government has left after paying for its spending
- public sector** that part of the economy where business activity is owned, financed and controlled by the state, and goods and services are provided by the state on behalf of the population as a whole
- purchasing power parity** a theory of exchange rates whereby a unit of any given currency should be able to buy the same quantity of goods in all countries
- quantity demanded** the amount of a good that buyers are willing and able to purchase at different prices
- quantity equation** the equation $M \times V = P \times Y$, which relates the quantity of money, the velocity of money, and the currency value of the economy's output of goods and services
- quantity supplied** the amount of a good that sellers are willing and able to sell at different prices
- quantity theory of money** a theory asserting that the quantity of money available determines the price level and that the growth rate in the quantity of money available determines the inflation rate
- random walk** the path of a variable whose changes are impossible to predict
- rational expectations** the theory according to which people optimally use all the information they have, including information about government policies, when forecasting the future
- rational ignorance effect** the tendency of a voter to not seek out information to make an informed choice in elections
- reaction function** the decision of one firm on a particular issue such as the profit maximizing output in response to the profit maximizing output decisions of its rivals
- real exchange rate** the rate at which a person can trade the goods and services of one country for the goods and services of another
- real GDP** the measure of the value of output in the economy which takes into account changes in prices over time
- real interest rate** the interest rate corrected for the effects of inflation
- real money balances** what money can actually buy given the ratio of the money supply to the price level $\frac{M}{P}$
- real variables** variables measured in physical units
- real wages** the money wage adjusted for inflation measured by the ratio of the wage rate to price $\frac{W}{P}$
- recession** a period of declining real incomes and rising unemployment. The technical definition gives recession occurring after two successive quarters of negative economic growth
- refinancing rate** the interest rate at which the European Central Bank lends on a short-term basis to the euro area banking sector
- regressive tax** a tax for which high-income taxpayers pay a smaller fraction of their income than do low-income taxpayers
- relative poverty** a situation where an individual is not able to access what would be considered acceptable standards of living in society
- rent seeking** where individuals or groups take actions to redirect resources to generate income (rents) for themselves or the group
- repo rate** the interest rate at which the Bank of England lends on a short-term basis to the UK banking sector
- repurchase agreement (repo)** the sale of a non-monetary asset together with an agreement to repurchase it at a set price at a specified future date
- reserve ratio** the fraction of deposits that banks hold as reserves
- reserve requirements** regulations on the minimum amount of reserves that banks must hold against deposits
- reserves** deposits that banks have received but have not loaned out
- residual demand** the difference between the market demand curve and the amount supplied by other firms in the market
- risk** the probability of something happening which results in a loss or some degree of hazard or damage
- risk averse** exhibiting a dislike of uncertainty
- rival** the property of a good whereby one person's use diminishes other people's use
- sacrifice ratio** the number of percentage points of annual output lost in the process of reducing inflation by 1 percentage point
- satisficers** those who make decisions based on securing a satisfactory rather than optimal outcome
- scarcity** the limited nature of society's resources
- screening** an action taken by an uninformed party to induce an informed party to reveal information
- shoeleather cost** the resources wasted when inflation encourages people to reduce their money holdings

shortage a situation in which quantity demanded is greater than quantity supplied at the going market price

signalling an action taken by an informed party to reveal private information to an uninformed party

Single European Market a (still not complete) EU-wide market throughout which labour, capital, goods and services can move freely

social security government benefits that supplement the incomes of the needy

social welfare function the collective utility of society which is reflected by consumer and producer surplus

special-interest effect where benefits to a minority special-interest group are outweighed by the costs imposed on the majority

specific tax a fixed rate tax levied on goods and services expressed as a sum per unit

stagflation a period of falling output and rising prices

standard of living refers to the amount of goods and services that can be purchased by the population of a country. Usually measured by the inflation-adjusted (real) income per head of the population

stationary data time series data that have a constant mean value over time

stochastic trend where trend variables change by some random amount in each time period

stock (or share or equity) a claim to partial ownership in a firm

store of value an item that people can use to transfer purchasing power from the present to the future

strike the organized withdrawal of labour from a firm by a union

structural deficit a situation where a government's deficit is not dependent on movements in the economic cycle

structural unemployment unemployment that results because the number of jobs available in some labour markets is insufficient to provide a job for everyone who wants one

sub-prime market individuals not traditionally seen as being part of the financial markets because of their high credit risk

subsidy payment to buyers and sellers to supplement income or lower costs and which thus encourages consumption or provides an advantage to the recipient

substitutes two goods for which an increase in the price of one leads to an increase in the demand for the other

substitution effect the change in consumption that results when a price change moves the consumer along a given indifference curve to a point with a new marginal rate of substitution

sunk cost a cost that has already been committed and cannot be recovered

supply curve a graph of the relationship between the price of a good and the quantity supplied

supply schedule a table that shows the relationship between the price of a good and the quantity supplied

supply shock an event that directly alters firms' costs and prices, shifting the economy's aggregate supply curve and thus the Phillips curve

surplus a situation in which the quantity supplied is greater than the quantity demanded at the going market price

synergies where the perceived benefits of the combined operations of a merged organization are greater than those which would arise if the firms stayed separate

tacit collusion when firm behaviour results in a market outcome that appears to be anti-competitive but has arisen because firms acknowledge that they are interdependent

tariff a tax on goods produced abroad and sold domestically

tax incidence the manner in which the burden of a tax is shared among participants in a market

technological knowledge society's understanding of the best ways to produce goods and services

the economy all the production and exchange activities that take place every day

the long run the period of time in which all factors of production can be altered

the short run the period of time in which some factors of production cannot be changed

theory of liquidity preference Keynes' theory that the interest rate adjusts to bring money supply and money demand into balance

time series data observations on a variable over a time-period and which are ordered over time

total expenditure the amount paid by buyers, computed as the price of the good times the quantity purchased

total revenue the amount received by sellers of a good, computed as the price of the good times the quantity sold

total surplus the total value to buyers of the goods, as measured by their willingness to pay, minus the total cost to sellers of providing those goods

total utility the satisfaction gained from the consumption of a good

toxic debt mortgage-backed securities and other debt (such as bonds) that are not able to be repaid in many cases because the value of the assets against which they are secured have fallen significantly

trade balance the value of a nation's exports minus the value of its imports; also called net exports

trade deficit an excess of imports over exports

trade policy a government policy that directly influences the quantity of goods and services that a country imports or exports

trade surplus an excess of exports over imports

Tragedy of the Commons a parable that illustrates why common resources get used more than is desirable from the standpoint of society as a whole

transaction cost the opportunity cost of carrying out a transaction in any market

transaction costs the costs that parties incur in the process of agreeing and following through on a bargain

transfer earnings the minimum payment required to keep a factor of production in its current use

transfer payment a payment for which no good or service is exchanged

trend the underlying long-term movement in a data series

trough where economic activity reaches a low and the decline ends

unemployment insurance a government programme that partially protects workers' incomes when they become unemployed

- unemployment rate** the percentage of the labour force that is unemployed
- union** a worker association that bargains with employers over wages and working conditions
- union density** a measure of the proportion of the workforce that is unionized
- unit of account** the yardstick people use to post prices and record debts
- utilitarianism** the political philosophy according to which the government should choose policies to maximize the total utility of everyone in society
- utility** the satisfaction derived from the consumption of a certain quantity of a product
- value of the marginal product** the marginal product of an input times the price of the output
- variable costs** costs that are dependent on the quantity of output produced
- velocity of money** the rate at which money changes hands
- vertical equity** the idea that taxpayers with a greater ability to pay taxes should pay larger amounts
- wealth** the total of all stores of value, including both money and non-monetary assets
- welfare economics** the study of how the allocation of resources affects economic well-being
- well-being** happiness or satisfaction with life as reported by individuals
- willingness to pay** the maximum amount that a buyer will pay for a good
- world price** the price of a good that prevails in the world market for that good

- ability-to-pay principle of taxation, 214–15
- abnormal profit, 155
- absolute advantage, 414–15
- absolute poverty, 393
- accelerator principle, 660
- accounting profit, 152–3
- actual spending, 656
- AD *see* aggregate demand
- adaptive expectations, 735–6
- ad valorem* taxes, 192, 194–6
- adverse selection, 266–8
- advertising, 48, 319–22, 339
- agents, 264–5
- aggregate demand, 669–75, 693–4
 - fiscal policy, 709–13
 - monetary policy, 702–9
 - Phillips curve, 722–4
 - taxation, 711–13
- aggregate demand curve, 681–5, 705–7
- aggregate risk, 546
- aggregate supply, 694–5, 722–4
- aggregate supply curve, 682, 686–91, 745–9
- airlines, 306
- allocative efficiency, 169–70
- amplitude, 641
- anchoring, 127
- anti-poverty policies, 400–1
- appreciation, 610
- arbitrage, 304, 613
- AS (aggregate supply), 694–5, 722–4
- asset valuation, 548–53
- assumption of rational behaviour, 246–7
- assumptions, 20
- asymmetric information, 246–7, 265–71, 275
- asymmetric shocks, 766
- austerity, 676–7, 797–801
- Austrian school of economics, 564
- automatic stabilizers, 715–16
- autonomous spending, 657
- availability heuristics, 128
- average costs, 140–6
- average tax rate, 212
- Axiom of Comparison, 108
- Axiom of Transitivity, 108

- balanced budgets, 716
- balanced trade, 606
- Banco Nacional de Desenvolvimento Econômico e Social, 536
- bandwagon effects, 128
- Bank of England, 567
- banks, 522–3, 568–79
 - central, 565–7, 730–2, 790–2
 - financial crisis, 576–9
- Barclays LIBOR scandal, 275
- bargaining, 246
- bar graphs, 32
- barriers to entry, 291–4
- barter, 558
- basketball, 143
- Becker's employer taste model, 376–7

- behavioural approaches to consumer behaviour, 126–9
- behavioural economics, 264–76
- benefits principle of taxation, 214
- bequests, 363
- Bertrand Model of oligopoly, 347
- Big Mac Index, 616–17
- BNDES (Banco Nacional de Desenvolvimento Econômico e Social), 536
- bond market, 520–1
- bonds, 520
- borrowing, 536
- bounded rationality, 126
- brain drain, 487
- branding, 322–3, 325–6
- Brazil, 536
- bubbles, 782–93
- budget constraints, 104–7
- bureaucrat incentives, 256
- business cycles, 11–12, 637–52
 - changes in, 645–6
 - confidence, 646
 - expectations, 646
 - external forces, 645
 - government policy, 646
 - models, 646–8
 - real, 648–50

- capital, 2
- capital flight, 632–3, 634–5
- capital gains tax, 596
- capital income, 378
- capital market, 377–8
- capital mobility, 766–7, 769
- capital stock, 491–2
- carbon price floors, 200
- carbon trading permits, 252–3
- Carnegie Conjecture, 363
- cars, 175, 267–8
- cartels, 331–2
- catch-up effect, 486
- cause and effect, 38–40
- CDOs (collateralized debt obligations), 526
- CDS (credit default swaps), 523–5, 787
- central banks, 565–7, 730–2, 790–2
- ceteris paribus, 35
- charities, 244
- China
 - comparative advantage, 417
 - demand, 82
 - economic growth, 493–5
 - inflation, 699
 - LCD cartels, 332
 - open market operations, 580
 - seafood, 82
- choice sets, 108
- Christmas, 115–16, 269
- cinema tickets, 306
- circular-flow diagrams, 22–3
- claimant counts, 498
- classical dichotomy, 587–9
- climate change, 235–6

- Coase theorem, 245–6
- coincident indicators, 643
- collateralized debt obligations, 526
- collective bargaining, 508
- collusion, 331, 350–1
- commodity money, 560–1
- common currencies, 762–8
- common currency areas, 614, 759, 772–7
- common resources, 222, 230–2, 235–6, 339–40
- comovement, 643
- comparative advantage, 414–18, 431–3
- Comparison, Axiom of, 108
- compensating differentials, 368
- competition, 6, 42–3
 - imperfect, 290–1
 - meaning of, 150
 - versus* monopoly, 294–5
 - product differentiation, 315–19
- competition law, 310–11, 348–50
- competitive advantage, 324
- competitive firms, 151–60
- competitive markets, 42–3, 150–3, 160–4
- complements, 47
- compounding, 540
- concentration ratio, 329
- constant returns to scale, 147
- constrained optimization, 40, 113–25
- consumer choice, 116–26
- consumer expectations, 48
- consumer prices index, 456–63
- consumer surplus, 170–5
- consumption, 443
- contestable markets, 323–4
- contingent valuation methods, 226
- contractions, 640–1
- coordinate system, 33–6
- copulas, 547
- copyright, 292–3
- corporate decision-making, 645
- correlation, 33
- cost-benefit analysis, 225–8
- cost curves, 141–4
- cost of living, 456–70
- costs
 - average, 140–6
 - of capital, 135
 - definitions, 146
 - explicit, 134–5
 - fixed, 139
 - implicit, 134–5
 - marginal, 140–1, 142–3, 165–6
 - measures of, 138–44
 - rising marginal, 142
 - sunk, 158
 - variable, 140
- cotton, 68–9
- countercyclical variables, 642–3
- coupons, 306
- Cournot Model of oligopoly, 345–6
- CPI (consumer prices index), 456–63
- cream-skimming, 324

- credit default swaps, 523–5, 787
 cronyism, 259
 cross-price elasticity of demand, 81–2, 93
 crowding out, 535, 536, 709–11
 currency, 561
 curves, 35–6
 customer preferences, 108–13
 cyclical deficit, 776, 797–8
 cyclical unemployment, 501
- dark pools, 527
 deadweight loss, 204–10, 213, 269, 301–2
 decision-making, 2–6, 28, 129–31, 645
 deductive reasoning, 18
 deflation, 599
 deflationary gap, 658
 demand, 43–9
 - cross-price elasticity of, 81–2
 - factors affecting, 46–8
 - income elasticity of, 81
 - law of, 43
 - price elasticity of, 72–81
 demand curves, 35–6, 43–8
 - consumer surplus, 171–2
 - derivation, 120–1
 - kinked, 334–5
 - linear, 80–1
 - movement along, 46, 48
 - shifts in, 46–7, 48
 - slope, 121–2
 - variety of, 76–7
 demand deposits, 562
 demand management, 658–9, 663–4
 demand schedules, 43–4
 demand-side policies, 647–8
 de-merit goods, 234
 depreciation, 610
 depressions, 637
 deregulation, 754–5, 783
 derived demand, 355
 destroyer pricing, 323
 deterministic trends, 642
 developing countries, 536
 dilution of capital stock, 491–2
 diminishing marginal product, 137–8
 diminishing marginal utility, 111
 diminishing returns, 485–6
 direct taxes, 192
 discount coupons, 306
 discount rate, 573
 discounts, 306
 discrimination, 373–6
 diseconomies of scale, 147–8
 disposable personal income, 442
 diversification, 544–7
 dominant strategy, 337–8
 double coincidence of wants, 558
 duopolies, 330
- ECB (European Central Bank), 566–7
 economic activity, 2–3
 economic activity rate, 499–500
 Economic and Monetary Union, 760, 768–72
 economic bubbles, 782–93
 economic efficiency, 179–81
 economic fluctuations, 679–82, 691–5
 economic growth, 9, 474–7, 478–9, 482–95, 800–1
- economic life cycles, 390–2
 economic mobility, 392–3
 economic problem, 1–2
 economic profit, 152–3
 economic rent, 379–80
 economics
 - meaning of, 1–2
 - relevance of, 431–3
 - 10 principles of, 3–11
 economic slowdown, 757
 economic well-being *see* well-being
 economies
 - equilibrium, 656–9
 - meaning of, 2–3
 - open, 605–19, 622–35
 - stabilization of, 713–16
 economies of scale, 147–9
 economists
 - agreement, 26–7
 - as decision makers, 28
 - disagreement, 24–5
 - as policy advisors, 23–4
 - scientific judgements, 24–5
 - as scientists, 17–23
 - values, 25
 education, 232–3, 369–70, 487, 755–6
 efficiency, 180
 efficiency wages, 372–3
 efficiency wage theory, 510–11
 efficient markets hypothesis, 549–56
 efficient scale, 142
 EFSF (European Financial Stability Fund), 796–7
 elasticity, 72–99
 - applications of, 95–7
 - of demand, 72–82
 - mathematics of, 90–5
 - oil prices, 98–9
 - of supply, 83–9
 - tax incidence, 196–7
 empiricism, 18–20
 employment, 651–2 *see also* unemployment
- EMU (European Economic and Monetary Union), 760, 768–72
 endogenous variables, 22
 endowment effect, 272–3
 Engel Curves, 124–5
 entry limit pricing, 323
 equilibrium in the goods market, 665
 equilibrium in the money market, 705
 equilibrium of the economy, 656–9
 equilibrium price, 56–7
 equilibrium quantity, 56–7
 equity, 3, 182–3, 214–16
 equity finance, 521
 ESM (European Stability Mechanism), 796–7
 Euro, 566, 760–2
 Euro-Budget, 778–9
 European Central Bank, 566–7
 European Economic and Monetary Union, 760, 768–72
 European Financial Stability Fund, 796–7
 European Stability Mechanism, 796–7
 European Union, 761
 excess capacity, 317–18
 exchange rates, 611–12, 618–19
 exchange rate variability, 763
 excludable goods, 222
 exits, 157, 158–9
- exogenous variables, 22
 expansionary austerity, 676–7
 expectations, 48, 646, 699
 expected utility theory, 128–9
 experiments, 20–1
 explicit costs, 134–5
 exports, 418, 607–8
 externalities, 8, 239–54
 - gun crime, 260–1
 - internalizing, 242–3
 - and market inefficiency, 240–4
 - negative, 239, 241–3
 - positional, 247–8
 - positive, 239, 243–4
 - private solutions, 244–8
 - public policies, 248–51
 - types of, 240
- factors of production, 378–9
 fairness, 273
 Fama, Eugene, 554–6
 fiat money, 561
 films, 466
 finance, 539
 financial capital mobility, 766–7, 769
 financial crisis, 525, 576–9, 782–802
 financial institutions, 520–6
 financial intermediaries, 522–3
 financial markets, 520
 financial system, 519–20
 fiscal compact, 776–7
 fiscal consolidation, 798–800
 fiscal federalism, 772–3
 fiscal multipliers, 717–18
 fiscal policy, 668, 709–13, 772–7
 Fisher effect, 592–3
 fishing, 231–2
 fixed costs, 139
 flat tax, 215
 flexible labour markets, 753–4
 foreign currency exchange, market for, 624–8
 foreign direct investment, 486–7
 foreign exchange rate variability, 763
 foreign portfolio investment, 486–7
 fracking, 55
 fractional-reserve banking, 568–9
 France, 381–2
 free riders, 223–4
 free trade, 490
 frictional unemployment, 503, 504–5
 full employment, 655–6
 functions, 30–1
 fundamental analysis, 548–9
 future value, 540
- game theory, 335–6
 GDP *see* gross domestic product
 general equilibrium, 179
 Gesell, Silvio, 710
 Giffen goods, 122
 Gini coefficient, 388–90, 391
 global recession, 789–90
 good governance, 488–9
 goods
 - excludable, 222
 - merit, 232–4
 - private, 222
 - public, 222, 223–30
 - rival, 222
 - types of, 222–3

- goods market, equilibrium in, 665
governance, 488–9
government budget deficits, 628–30
government budgets, 534–5
government failure, 254–9
government policy, 187–200, 646
see also public policy
government purchases, 443, 709
government spending reduction, 754
graphs, 30–40
single variable, 31–3
two variables, 33–40
types of, 31–3
Greece, 793–4
greenhouse gas emissions, 235–6
gross domestic product, 438–52
components of, 441–5, 528–9
at constant prices, 445–6
at current prices, 445–6
deflator, 447–8, 463
at market prices, 445–6
measurement, 439–41, 448
nominal, 446–7
per capita, 9–10, 445
real, 445–7
revisions, 444
and well-being, 449–52
gross investment, 479
gross national product, 442
growth *see* economic growth
Gulf Cooperation Council states, 310–11
gun crime, 260–1
- happiness, 449–51
harmonized indices of consumer prices, 463
health, 487–8, 510
healthcare, 233–4
hedonics, 469–70
herd mentality, 272
heuristics, 127–8
HICPs (harmonized indices of consumer prices), 463
horizontal equity, 214, 215
household spending, 645
house prices, 783–5
human capital, 368–70, 480
human life, valuation of, 226–8
hysteresis, 502
- ice cream, 583, 598
Iceland, 696
idiosyncratic risk, 544–6
imperfect competition, 290–1
implicit costs, 134–5
import quotas, 425–7
imports, 418
incentives, 5–6, 12–14
income
consumer choice, 116–17
demand curve, 47–8
inequality, 386–93, 402
redistribution of, 395–401
income effects, 118–19
income elasticity of demand, 81, 92–3
income expansion path, 122–4
inconsistency, 274
indexation, 465–6
India, 350–1
indicators, 643
indifference curves, 108–10
- indirect taxes, 192
inductive reasoning, 18
industrial policies, 756
inequality, 386–93, 402
inferior goods, 48
inflation, 11, 583–98
China, 699
correction for, 465–7
cost of, 594–9
cost of reduction, 734–7
fear of, 726
and money supply, 600–2
targeting, 739–40, 741–2
theory of, 584–93
unemployment, 721–42
inflationary gap, 658
inflation rate, 458
inflation tax, 592
informational efficiency, 550
infrastructure investment, 755
in-kind transfers, 400
insurance, 233–4, 268, 542–3
interaction, 6–8
interest income, 596
interest rates, 467, 706, 709–10
internalizing an externality, 242–3
international trade, 410–14, 429
intertemporal choice, 232
interventionist supply-side policies, 749, 755–6
introduction of new goods, 459
investment, 443, 483–4, 527–30, 608–9
in education and training, 755–6
incentives, 533
in infrastructure, 755
investment funds, 523
Ireland, 525
IS (investment and saving) curve, 665–6
IS-LM model, 665, 667–9, 671–2
IS-MP model, 672–3
isocost lines, 282–4
isoquants, 279–82
- job search, 504–6
- Keynes, John Maynard, 553, 647, 655
Keynesian cross, 655–9
Keynesian economics, new, 697–8
Keynesian intervention, 673–5, 676–7
kinked demand curves, 334–5
Krugman, Paul, 674–5
Kydland, Finn E. 649–50
- labour, 2
labour demand, 356–60, 365
labour force, 499
labour force participation rate, 499–500
labour force surveys, 499–500
labour markets, 268, 355–82, 753–4
labour mobility, 766, 769
labour productivity, 286
labour supply, 360–4, 365
labour unions, 371–2
Laffer curve, 752–3
lagging indicators, 643
land, 2
land market, 377–8
law of demand, 43
law of supply, 50
law of supply and demand, 58
LCD (liquid crystal display panels), 332
- leading indicators, 643
least-cost input combination, 284–6
Lehman Bros, 789
Li, David, 547–8
liberalism, 396–7
libertarianism, 397–8
libertarian paternalism, 398
LIBOR (London Interbank Offered Rate), 275, 577–8
life cycles, 390–2
linear demand curves, 80–1
linear equations, 31
line graphs, 33
liquid crystal display panels, 332
liquidity, 560
liquidity preference, theory of, 703–5
LM (liquidity and money) curve, 666–7
loanable funds, market for, 530–5, 623–4, 626–8
logrolling, 257–8
London Interbank Offered Rate, 577–8
long run, 136
long-term unemployment, 502
Lorenz curve, 386–8
lotteries, 363
lump-sum tax, 212
- macroeconomics, 9, 437–53, 605–19, 622–35
Malthus, Thomas Robert, 491
marginal abatement cost, 235, 249
marginal changes, 5
marginal cost curves, 155–7
marginal costs, 140–1, 142–3, 165–6
marginal product, 136–8
marginal product of labour, 356–9
marginal propensity to consume, 660–1
marginal propensity to save, 660–1
marginal rate of substitution, 111–12
marginal rate of technical substitution, 281
marginal revenue, 151–2
marginal revenue product, 358
marginal tax rate, 212
marginal utility, 110–11
market demand, 44–6
market economy, 6–8
market efficiency, 179–83
market entry, 158–9
market equilibrium, 59–62, 62–7, 181–2
market exit, 157, 158–9
market failure, 8
market for foreign currency exchange, 624–8
market for loanable funds, 530–5, 623–4, 626–8
market inefficiency, 240–4
market irrationality, 552–3
market-orientated supply-side policies, 749–55
market power, 8, 290–1
market price, 56–7
markets, 41–3
market segments, 330
market share, 290–1
market supply, 50–1
maximin criterion, 396–7
mediums of exchange, 559
mental accounting, 271–2
menu costs, 595–6
merit goods, 232–4

- microeconomics, 9
- minimum efficient scale, 142
- minimum wage, 26, 370–1, 398–9, 506–7
- Minsky's financial instability hypothesis, 552
- misallocation of resources, 596
- Mises, Ludwig von, 564
- misperceptions theory, 690
- model of aggregate demand and aggregate supply, 681–2
- models, 21–3
- monetary control, 571–3
- monetary equilibrium, 584–7
- monetary injection, 586–7
- monetary neutrality, 588–9
- monetary policy, 565, 618–19, 669, 702–9
- money
 - meaning of, 559–62
 - quantity theory of, 587
 - value of, 584
 - velocity of, 589–90
- money demand, 584–6, 704–5
- money markets, 572, 705
- money multiplier, 570–1
- money stock, 561
- money supply, 564, 565, 575, 584–6, 600–2, 703–4, 707–8
- monopolies, 43, 290–311
 - causes of, 291–4
 - versus* competition, 294–5
 - competition law, 310–11
 - deadweight loss, 301–2
 - government-created, 292–3
 - natural, 293–4
 - pharmaceuticals, 299–300
 - prevalence of, 309–10
 - profit, 298
 - profit maximization, 297–8
 - public ownership, 309
 - public policy, 307–9
 - regulation, 308–9
 - revenue, 295–7
 - social cost, 302–3
 - supply curves, 297–8
 - welfare cost, 300–3
- monopolistically competitive markets, 43
- monopolistic competition, 314–26
- monopoly resources, 292
- monopsony, 366–7
- moral hazard, 265–6, 275
- MP (monetary policy) function, 672–3
- multiplier effect, 659–64

- naïve empiricism, 127
- Nash equilibrium, 333, 340–2
- national defence, 224
- national income, 442
- national savings, 528–9
- natural experiments, 21
- natural monopolies, 222, 293–4
- natural-rate hypothesis, 729
- natural rate of output, 687–8
- natural rate of unemployment, 501–2, 726
- natural resources, 480, 482
- negative correlation, 33
- negative externalities, 239, 241–3
- negative income tax, 399–400
- negative interest rates, 710
- net capital outflow, 606–9, 625–6
- net exports, 444–5, 607–8
- net investment, 479
- net national product, 442
- neuroscience, 129–31
- New Keynesian economics, 697–8
- nominal exchange rates, 610
- nominal gross domestic product, 446–7
- nominal interest rates, 467
- nominal values, 40
- nominal variables, 587–8
- non-price competition, 335
- nonstationary data, 641
- non-tariff barriers, 427
- normal goods, 47–8
- normal profit, 155
- normative statements, 23–4
- nutrition, 487–8

- oil, 339–40
- oil prices, 55, 98–9
- Okun's law, 680
- oligopolies, 43, 329–51
 - characteristics of, 329–35
 - models of, 344–7
 - prisoners' dilemma, 338–9
 - product differentiation, 330
 - public policy, 347–50
- omitted variables, 38–9
- open economy macroeconomics, 605–19, 622–35
- open-market operations, 565, 571, 580
- opportunity costs, 4, 135, 407, 415
- optimum currency areas, 765–8, 768–72
- output, 680
 - natural rate of, 687–8
- output gap, 658
- outright open-market operations, 571
- overconfidence, 126

- Pakistan, 175
- Pareto improvement, 180–1
- participation rate, 499–500
- patents, 292–3
- payoff matrix, 336
- peaks, 640
- pensions, 233–4
- perfect complements, 112–13
- perfect price discrimination, 305
- perfect substitutes, 112
- permanent income, 392
- personal income, 442
- persuasion heuristics, 128
- pharmaceuticals, 299–300
- Phillips curve, 11, 721–34
- photocopying, 350–1
- physical capital, 479–80
- pie charts, 32
- Pigovian taxes, 248–9
- PIIGS (Portugal, Ireland, Italy, Greece and Spain) countries, 771, 773, 795, 802
- planned spending, 656
- political stability, 488–9
- politician incentives, 256
- pollution, 231, 235–6
 - tradable permits, 249–51, 252–4
- population, 48
- population growth, 491–3
- positional arms races, 248, 253
- positional externalities, 247–8
- positive correlation, 33
- positive externalities, 239, 243–4
- positive statements, 23–4
- poverty, 225
- poverty line, 393
- poverty rate, 393
- poverty reduction, 398, 400–1
- PPS (purchasing power standard), 615
- predatory pricing, 323, 349
- Prescott, Edward C. 649–50
- present value, 539–41
- price ceilings, 187–90
- price-consumption curves, 120
- price controls, 187–92
- price discrimination, 303–6, 762–3
- price elasticity of demand
 - applications of, 95–7
 - arc method of calculation, 74–5
 - computing, 73–6
 - linear demand curves, 80–1
 - mathematics of, 90–1
 - midpoint method of calculation, 74–5
 - point method of calculation, 75–6
 - rules, 73
 - total expenditure, 78–81, 94–5
- price elasticity of supply
 - applications of, 95–7
 - arc method of calculation, 85
 - computing, 85
 - determinants of, 83–4
 - mathematics of, 93–4
 - midpoint method of calculation, 85
 - point method of calculation, 85–6
 - total revenue, 88–9, 94–5
- price floors, 187, 190–2, 200
- price indices, 456–63, 469–70
- price levels, 681
- prices
 - consumer choice, 117–18
 - consumer surplus, 173
 - finding, 49
 - producer surplus, 178–9
 - resource allocation, 67
 - as signals, 58–9
- principals, 264–5
- prisoners' dilemma, 337–44
- private goods, 222
- private savings, 529
- private sector, 221
- private solutions to externalities, 244–8
- privatization, 258–9, 754–5
- procyclical variables, 642–3
- producer prices index, 458
- producer surplus, 176–9
- product differentiation, 315–19, 330
- production costs, 134–8
- production decisions, 287–8
- production function, 136–8
- production isoquants, 279–82
- production possibilities frontier, 405–10
- productivity, 10–11, 286, 477–81
- profit maximization, 153–4, 297–8
- profit measurement, 159–60
- progressive tax, 215
- property rights, 251–2, 488–9
- proportional tax, 215
- prospect theory, 272
- public choice theory, 255
- public goods, 222, 223–30
- public interest, 255–6

- public policy *see also* government policy
 asymmetric information, 270–1
 economic growth, 482–93
 externalities, 248–51
 job search, 505
 monopolies, 307–9
 oligopolies, 347–50
 public savings, 529
 public sector, 221
 purchasing power, 594–5
 purchasing power parity, 612–17, 626
 purchasing power standard, 615
- Qatar, 310–11
 QE (quantitative easing), 574–5, 600–2, 618–19
 quality of life, 451–2
 quantitative easing, 574–5, 600–2, 618–19
 quantity demanded, 43, 49
 quantity discounts, 306
 quantity equation, 590
 quantity supplied, 50
 quantity theory of money, 587
 quants, 547–8
 quotas, 25–6, 425–7
- random walks, 550
 RAS (reticular activation system), 127
 rational behaviour, assumption of, 246–7
 rational expectations, 735
 rational ignorance effect, 256
 rationalism, 19–20
 rationality, 126–9
 reaction function, 346
 real business cycles, 648–50
 real exchange rates, 611–12
 real gross domestic product, 445–7
 real interest rates, 467
 real money balances, 665
 real values, 40
 real variables, 587–8
 real wage flexibility, 766, 768–9
 real wages, 588
 recessions, 452–3, 637
 redistribution of income, 395–401
 redistribution of wealth, arbitrary, 598
 refinancing rate, 571–3
 regional policies, 756
 regressive tax, 215
 regulation, 251, 308–9
 relative poverty, 393
 relative price variability, 596
 rent control, 189–90
 rent seeking, 258
 repo rate, 572
 representativeness heuristics, 128
 repurchase agreements, 572
 resale price maintenance, 348–9
 research, 224–5
 research and development, 490, 756
 reserve ratio, 569
 reserve requirements, 573
 reserves, 568
 residual demand, 345
 resource allocation, 67
 resources, misallocation of, 596
 restraint of trade, 348
 retail prices index, 461–2
 reticular activation system (RAS), 127
- returns to scale, 147
 revenue, competitive firms, 151–2
 reverse causality, 39–40
 Ricardo, David, 416
 rising marginal costs, 142
 risk aversion, 541–2
 risk management, 541–8
 rival goods, 222
 road congestion, 231, 255
 Romer model, 672–3
 RPI (retail prices index), 461–2
 rule of 70, 476
 Russia, 634–5
- sacrifice ratio, 734–5
 Saha, David, 674
 satisficers, 271
 savings, 483–4, 527–33, 608–9
 scarcity, 2
 scatterplots, 33–5
 scientific method, 19
 screening, 270
 seafood, 82
 securitization of assets, 576–8
 self-interest, 244
 SEM (standard economic model), 102–3, 126, 271–4
 SGP (Stability and Growth Pact), 775
 shares, 521
 ships, 287–8
 shoelace costs, 595
 shortages, 57–8
 short run, 136
 short-termism, 258
 shutdowns, 157
 signalling, 268–70
 simulation heuristics, 128
 single currency, 762–8
 Single European Market, 761–2
 slope, 37–8
 demand curves, 121–2
 supply curves, 164
 Smith, Adam, 7, 416
 social contracts, 244–5
 socially efficient outcomes, 242–3
 social norms of moral behaviour, 244
 social optimum, 242–3
 social security, 399
 social welfare function, 183
 South Sea Bubble, 786
 sovereign debt crisis, 793–7, 802
 special-interest effect, 257
 special purpose vehicles, 576–8
 specific taxes, 192–4
 spending, 656–7
 SPV (special purpose vehicles), 576–8
 Stability and Growth Pact, 775
 stabilization of economies, 713–16
 Stackelberg Model of oligopoly, 347
 stagflation, 695–6
 standard economic model, 102–3, 126, 271–4
 standard of living, 10–11
 stationary data, 641
 sticky price theory, 690
 sticky wage theory, 689–90
 stochastic trends, 642
 stock markets, 521–2, 553
 stocks, 521
 stores of value, 560
 strikes, 372
- structural deficit, 776, 797–8
 structural unemployment, 503
 sub-prime market, 526, 785–7
 subsidies, 198–9
 substitutes, 47
 substitution bias, 459
 substitution effects, 118–19
 sunk costs, 158
 superstars, 370
 supply, 50–6, 83–9
 supply and demand, 41, 56–67
 supply curves
 competitive firms, 153–60
 competitive markets, 160–4
 factors affecting, 53–4
 market supply, 50–1
 monopolies, 297–8
 price, 50
 producer surplus, 177
 quantity supplied, 50
 shifts in, 51–4, 56
 slope, 164
 variety of, 86
 supply schedules, 50
 supply shocks, 732–4
 supply-side policies, 646–7, 745–57
 economic slowdown, 757
 interventionist, 749, 755–6
 market-orientated, 749–55
 types of, 749–56
 surpluses, 57–8
 symmetric demand shocks, 767, 770–2
 synergies, 307
- tacit collusion, 344
 tariffs, 25–6, 424–5
 tastes, 48
 taxation
 ability-to-pay principle, 214–15
 administrative burden, 210–11
 ad valorem, 192, 194–6
 aggregate demand, 711–13
 average rates, 212
 benefits principle, 214
 canons of, 211
 capital gains tax, 596
 deadweight loss, 204–10, 213
 direct, 192
 and efficiency, 203–4
 and equity, 214–16
 flat, 215
 indirect, 192
 inefficiencies, 259
 interest income, 596
 lump-sum, 212
 marginal rates, 212
 Pigovian, 248–9
 principles of, 211
 progressive, 215
 proportional, 215
 reform, 750–1
 regressive, 215
 specific, 192–4
 system design, 211–12
 and welfare, 204–7
 tax avoidance, 217–18
 tax distortions, 596–7
 tax incidence, 192–3, 196–7, 215–16
 Taylor rule, 712
 technological knowledge, 480–1
 technological progress, 492–3

- 10 principles of economics, 3–11
- text messaging, 165–6
- theories, 18–19
- theory of liquidity preference, 703–5
- threshold concepts, 17
- time series data, 638
- time-series graphs, 32
- tools, 30–40
- total cost curves, 138
- total expenditure, 78–81, 94–5
- total revenue, 88–9, 94–5
- total surplus, 180
- total utility, 110–11
- toxic debt, 578
- tradable pollution permits, 249–51, 252–4
- trade, 6, 405–33
 - determinants of, 418–20
 - restrictions on, 423–30
 - winners and losers, 420–3
- trade balance, 606
- trade deficit, 606
- trade integration, 767–8
- trade-offs, 3–4, 546–7
- trade policy, 630–2
- trade surplus, 606
- trade unions, 508–9
- Tragedy of the Commons, 230–1
- training, 755–6
- transaction costs, 246, 595, 762
- transfer earnings, 379
- transfer payments, 443
- Transitivity, Axiom of, 108
- trend growth, 476
- trend growth rates, 638–43
- trends, 641–2
- troughs, 640
- tying, 349–50
- unemployment, 497–515, 651–2, 680
 - cost of, 511–13
 - cyclical, 501
 - frictional, 503, 504–5
 - and inflation, 721–42
 - long-term, 502
 - natural rate of, 501–2, 726
 - structural, 503
 - and vacancies, 514–16
- unemployment-inflation trade-off, 728–30
- unemployment insurance, 505–6
- unemployment rate, 499–500
- union density, 508
- unions, 371–2, 508–9
- units of account, 559–60
- unmeasured quality change, 459–61
- used cars, 267–8
- U-shaped average total cost, 142
- utilitarianism, 395–6
- utility, 103
- vacancies, 514–16
- value, 102–3
 - stores of, 560
- value of money, 584
- value of the marginal product, 358
- variable costs, 140
- velocity of money, 589–91
- vertical equity, 214–15
- very large ships, 287–8
- voter incentives, 256
- wage curve theory, 737–9
- wage differentials, 367–73
- wage flexibility, 766, 768–9
- waste, 179–80
- wealth, 560
- Weizsäcker, Jakob von, 674
- welfare, 204–7, 318–19 *see also*
 - well-being
- welfare economics, 169, 241
- welfare policy, 750–1
- well-being, 169, 184–5, 449–52 *see also*
 - welfare
- whaling, 231–2
- wildlife conservation, 231–2
- willingness to pay, 170–1
- willingness to sell, 176
- worker effort, 510–11
- worker health, 510
- worker quality, 511
- worker turnover, 511–13
- work incentives, 400–1
- world prices, 419–20
- Xerox stores, 350–1
- zero risk, 547–8

All Figures, Tables and artwork *not* listed on this credit page are the authors' own work and so do not require any credit lines, permissions acknowledgements or referencing citations.

IMAGES

The following Images have been reproduced with the kind permission of the copyright holders:

- Image 1.1 A typical supermarket shelf ... p. 4.
Source: © Kumar Sriskandan/Alamy
- Image 1.2 Adam Smith, p. 7. Source: © Classic Image/Alamy
- Image 1.3 Providing a monetary incentive ... p. 13.
Source: © Shutterstock
- Image 2.1 Models are used in many different disciplines ... p. 21. Source: © Shutterstock
- Image 3.1 The fracking process ... p. 55. Source: © Shutterstock
- Image 3.2 Changes in the price of cotton ... p. 69.
Source: © Andrew Ashwin
- Image 4.1 Fish farms like this in China ... p. 82.
Source: © Shutterstock
- Image 4.2 As supplies of oil start to diminish ... p. 98.
Source: © Shutterstock
- Image 5.A Would you pay €75 for these shoes ... p. 103.
Source: © Shutterstock
- Image 5.1 Many stores provide information on their receipt ... p. 116. Source: © Shutterstock
- Image 5.2 The brain plays a crucial role in decision making ... p. 131. Source: © Shutterstock
- Image 6.1 Thinking about average and marginal costs ... p. 143. Source: © Shutterstock
- Image 6.1a Thinking at the margin means that farmers ... p. 154. Source: © Shutterstock
- Image 6.2 With 1/1000th of a cent being the average cost ... p. 165. Source: © Lana Rastro/Alamy
- Image 7.1 The car in Pakistan ... p. 175. Source: © Asianet-Pakistan/Alamy
- Image 7.2 Poverty and wealth often cohabit side-by-side ... p. 184. Source: © Shutterstock
- Image 8.1 Taxes on the sale of goods and services ... p. 192. Source: © theodore liasi/Alamy
- Image 8.2 The carbon price floor ... p. 200. Source: © Shutterstock
- Image 9.1 The size of the deadweight loss ... p. 213.
Source: © Shutterstock
- Image 9.2 Amazon, Google and Starbucks attracted a good deal of publicity ... p. 218. Sources: © Marek Stepan/Alamy (left); © Steve Stock/Alamy (center); and © Biz/ed (right)
- Image 10.1 Accidents will happen ... p. 228. Source: © Shutterstock
- Image 10.2 The dumping of potentially hazardous waste ... p. 236. Source: © Shutterstock
- Image 11.1 How does a village become carbon neutral ... p. 253. Source: © Dave Ellison/Alamy
- Images 11.2a and 11.2b The costs of gun crime can be substantial ... p. 261. Source: © Shutterstock
- Image 12.1 Giving gifts is fraught with difficulty ... p. 269.
Source: © Caro/Alamy
- Image 12.2 Joseph Stiglitz won the Nobel Peace Prize ... p. 275. Source: © Gallo Images/Alamy
- Image 13.1 Levels of education have a significant impact on labour ... p. 286. Source: © Shutterstock
- Image 13.2 Very large ships help reduce unit costs ... p. 287. Source: © Shutterstock
- Image 14.1 Pfizer held the patent for the drug Viagra until July 2013 ... p. 300. Source: © Shutterstock
- Image 14.2 Where competition law and regulation is weak ... p. 311. Source: © Shutterstock
- Image 15.1 Could you tell the difference between Pepsi and Coca-Cola ... p. 320. Source: © Lux Igitur/Alamy
- Image 15.2 Staples Office Superstores ... p. 326.
Source: © Ian Dagnall/Alamy
- Image 16.1 Go into any electronics retailer ... p. 332.
Source: © Shutterstock
- Image 16.2 Photocopying shops in one part of India ... p. 351. Source: © Peter Horree/Alamy
- Images 17.1a and 17.1b If you won the lottery ... p. 363.
Source: © Shutterstock
- Image 17.2 An empty platform at the Gare de l'Est railway station ... p. 381. Source: © Shutterstock
- Image 18.1 Countries with relatively high Gini coefficients ... p. 391. Source: © Shutterstock
- Image 18.2 Family relationship problems can have ... p. 402. Source: © Shutterstock
- Image 19.1 A country's comparative advantage changes over time ... p. 417. Source: © Henry Westheim Photography/Alamy
- Image 19.2 Lee Kuan Yew, the first prime minister of independent Singapore ... p. 433. Source: © Asia File/Alamy
- Image 20.1 The complexity of capturing data ... p. 444.
Source: © Shutterstock
- Image 20.2 Observe the similarities between this 1930s Depression ... p. 453. Sources: © Everett Collection Historical/Alamy (left) and © Geoffrey Robinson/Alamy (right)
- Image 21.1 The number one film of all time ... p. 466.
Source: © Robert Harding Picture Library Ltd/Alamy
- Image 21.2 Technology changes over time ... p. 470.
Source: © Shutterstock
- Image 22.1 Copper prices fluctuate over short periods of time ... p. 482. Source: © Shutterstock

- Image 22.2 Professor Justin Yifu Lin ... p. 494. *Source:* © Justin Yifu Lin
- Image 22.3 Thomas Robert Malthus ... p. 491 © The Print Collection/Alamy
- Image 23.1 The rise in youth unemployment ... p. 502. *Source:* © Shutterstock
- Image 23.2 In reality the labour market ... p. 515. *Source:* © Shutterstock
- Image 24.1 Austerity policies have affected ... p. 525. *Source:* © Shutterstock
- Image 24.2 An offshore oil rig ... p. 536. *Source:* © Shutterstock
- Image 25.1 David Li was referred to by some ... p. 548. *Source:* © Ken Tam/Alamy
- Image 25.2 Eugene Fama was acknowledged ... p. 555. *Source:* © Eugene Fama, 2012
- Image 26.1 Ludwig von Mises ... p. 564. *Source:* © Ludwig von Mises
- Image 26.2 The People's Bank of China ... p. 580. *Source:* © TAO Images Limited/Alamy
- Image 27.1 Some question whether quantitative easing ... p. 601. *Source:* © Copyright 2007-2013 Truth Alliance inc.
- Image 28.1 Wherever you go in the world the Big Mac ... p. 617. *Source:* © incamerastock/Alamy
- Image 28.2 The Bank of England ... p. 619. *Source:* © Biz/ed
- Image 29.1 Money from criminal gangs ... p. 634. *Source:* © Shutterstock
- Image 30.1 Finn E. Kydland (left) and Edward C. Prescott (right) ... p. 650. *Source:* © WOLFGANG RATTAY/Reuters/Corbis
- Image 30.2 Productivity is a measure of output per worker ... p. 652. *Source:* © Shutterstock
- Image 31.1 To produce goods ... p. 660. *Source:* © MOHAMED ABD EL GHANY/Reuters/Corbis
- Image 31.2 Robert Skidelsky, Emeritus Professor ... p. 677. *Source:* © Robert Skidelsky
- Image 32.1 The collapse of the banking system ... p. 696. *Sources:* © Shutterstock (left) and © Ecojoe/Alamy (right)
- Image 32.2 New housing development in China ... p. 699. *Source:* © Shutterstock
- Image 33.1 Former Bank of England deputy governor ... p. 710. *Source:* © epa european pressphoto agency b.v./Alamy
- Image 33.2 Large-scale investment ... p. 718. *Source:* © epa european pressphoto agency b.v./Alamy
- Image 34.1 Rising fuel prices ... p. 726. *Source:* © Shutterstock
- Image 34.2 Mervyn King ... p. 742. *Source:* © epa european pressphoto agency b.v./Alamy
- Image 35.1 Arthur Laffer ... p. 753. *Source:* © Bob Daemmrlich/Cross/Alamy
- Image 35.2 Reform of planning laws ... p. 757. *Source:* © Shutterstock
- Image 36.1 Critics of the SGP ... p. 775. *Source:* © Cengage Learning EMEA
- Image 36.2 There are stark contrasts in economic welfare across Europe ... p. 779. *Source:* © Shutterstock
- Image 37.1 *The South Sea Scheme* ... p. 786. *Source:* © Falkensteinfoto/Alamy
- Image 37.2 One of the major fears of the problems ... p. 802. *Source:* © Shutterstock

FIGURES

The following Figures have been reproduced with the kind permission of the copyright holders:

- Figure 18.5 *Gini coefficient (measured on a scale from 0–100) for the EU 27 countries, 2011*, p. 390. *Source:* Eurostat. Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.
- Figure 18.8 *People at Risk of Poverty and Social Exclusion in the EU in 2011*, p. 394. *Source:* Eurostat. Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.
- Figure 21.1 *Two Measures of Inflation*, p. 464. *Source:* ONS and HM Treasury, Crown Copyright and public domain.
- Figure 21.2 *GDP Deflator (EU 27)*, p. 464. *Source:* Eurostat. Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.
- Figure 21.3 *Real and Nominal Interest Rates*, p. 468. *Source:* Bank of England and UK Office for National Statistics.
- Figure 22.1 *EU 27 GDP at Market Prices, Purchasing Power Standard Per Capita, 2003–2011 (€)*, p. 474. *Source:* Eurostat. Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.
- Figure 22.2 *Growth and Investment*, p. 484. *Source:* Fair use as author calculations and some World Bank data.
- Figure 22.4 *Corruption Perceptions Index 2012*, p. 489. *Source:* CPI Transparency
- Figure 23.1 *UK Unemployment Rate Since 1971*, p. 501. *Source:* UK Office for National Statistics and HM Treasury. Crown Copyright and public domain.
- Figure 26.1 *Three Measures of the Money Stock for the Euro Area*, p. 563. *Source:* ECB
- Figure 27.3 *The Velocity of Circulation, 1959–2013*, p. 591. *Source:* Economagic.org
- Figure 27.4 *The UK Nominal Interest Rate and the Inflation Rate*, p. 594. *Source:* Bank of England and UK Office for National Statistics. © Bank of England and public domain.
- Figure 30.1 *GDP in the UK and Europe, 1960–2011*, p. 639. *Source:* World Bank

Figure 30.2 *UK GDP Growth Rate, 1960–2011 (%)*, p. 640.

Source: World Bank

Figure 30.3 *OECD Area Composite Leading Indicators*

(CLI) Index, 2003–2013, p. 644. Source: OECD

Figure 36.3 *Growth Rates in Germany, France and*

the Whole Euro area, p. 770. Source: Eurostat.

Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.

Figure 36.4 *Growth Rates in Ireland, Greece and*

the Whole Euro area, p. 771. Source: Eurostat.

Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.

Figure 37.1 *The Rise in House Prices*, p. 784. Source:

Halifax House Price Index and Standard & Poors © Lloyds Banking Group and © 2013 Standard & Poor's Financial Services LLC, a part of McGraw Hill Financial. All rights reserved.

TABLES

The following Tables have been reproduced with the kind permission of the copyright holders:

Table 1.1, p. 10. Source: Fair use, data compiled from a variety of sources including <http://stats.oecd.org/index> and <http://data.worldbank.org/indicator>

Table 1.2, p. 12. Source: This is a blog which has been confirmed to be in the public domain:

<http://signsofchaos.blogspot.com/2005/11/price-elasticity-of-supply-and-web.html>

Table 2.1, p. 26. Source: Adapted from Richard M. Alston, J.R. Kearn and Michael B. Vaughn, 'Is There Consensus among Economists in the 1990s?' *American Economic Review* (May 1992): 203–209. Used by permission.

Table 4.2, p. 84. Source: Fair use, data compiled from a variety of sources including World Bank and UNESCO.

Table 21.2, p. 460. Source: Office of National Statistics, public domain.

Table 21.3, p. 462. Source: Eurostat. Reproduction of Eurostat data/documents for both commercial and non-commercial dissemination are automatically authorised provided appropriate acknowledgement is given to Eurostat as the source.

Table 21.4, p. 466. Source: The Movie Times (www.the-movie-times.com)

Table 22.1, p. 475. Source: World Bank

Table 23.1, p. 500. Source: Office of National Statistics, public domain.

Table 35.1, p. 750. Source: OECD

CASE STUDIES

No Case Studies text required permission.

IN THE NEWS

The following In The News items have been reproduced with the kind permission of the copyright holders:

Ch 1 In The News *Latest Thinking in Economics: Incentives*, p. 12. Source: Adapted from Gneezy, U. Meier, S & Ray-Biel, p. (2011). 'When and why incentives (don't) work to modify behaviour'. In *Journal of Economic Perspectives*. 25:4, pp. 191–210.

Ch 4 In The News *The Future Price of Oil and Elasticity*, p. 97. Source: www.washingtonpost.com by Brad Plumer, October 2012, © 1996–2013. The Washington Post.

Ch 10 In The News *The Problem of the Commons: Still Unsettled after 100 Years*, p. 235. Source: Permission not needed as fair use but references used include Stavins, Robert N, (2011). *The Problem of the Commons: Still Unsettled after 100 Years*. In *The American Economic Review*, 101, (1), pp. 81–108.

Ch 12 In The News *Moral Hazard*, p. 275. Source: 'None of them have gone to jail': man who ran World Bank calls for bankers to face the music'. *The Independent* (London, England) July 2nd 2012 © independent.co.uk

Ch 13 In The News *Production Decisions and the economics of v big ships*, p. 287. Source: 'Economies of scale made steel: Shipping', *The Economist* 12th November 2011. Copyright © The Economist Newspaper Limited 2013.

Ch 15 In The News *Best Brands are Disruptively Consistent*, p. 325. Source: *Marketing Week*, June 28, 2012, © 2012 Centaur Communications Limited.

Ch 18 In The News *Concerns Over Rate of Young Male Suicides*, p. 402. Source: www.irishexaminer.com January 24th 2013 © Irish Examiner Ltd, City Quarter, Lapps Quay, Cork.

Ch 19 In The News *Time to Review Economics Curriculum*, p. 431. Source: www.businessdayonline.com 29th March 2012, Ejike Nwolisa, © Business Day Online, 2013.

Ch 36 In The News *Does The Eurozone Need a Budget?* p. 778. Source: www.euractiv.com 22nd March 2013, Daniel Gros © EurActiv.com PLC, 2013.

FYIS

No FYI text required permission.