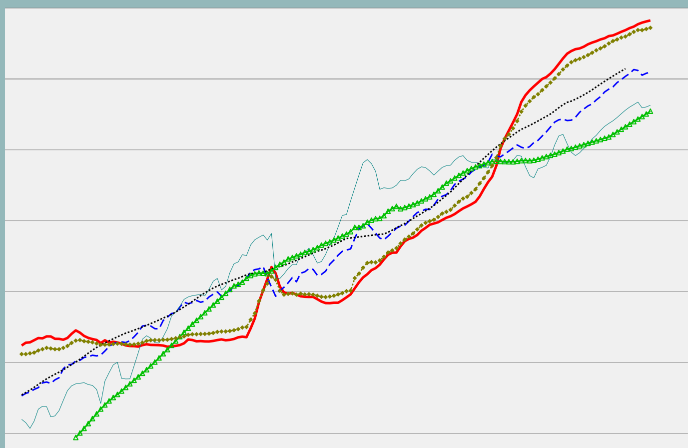
Introductory Macro-econometrics: A New Approach

David F. Hendry



Published by Timberlake Consultants Ltd
www.timberlake.co.uk
www.timberlake-consultancy.com

Introductory Macro-econometrics:
A New Approach

First published by Timberlake Consultants in 2015

British Library Cataloguing-in-Publication Data
A catalogue record of this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data
A catalogue record of this book is available from the Library of Congress
David F. Hendry
p. cm. – (Introductory Macro-econometrics:
A New Approach)

ISBN 978-0-9932738-0-3

Published by
Timberlake Consultants Ltd
Unit B3, Broomsleigh Business Park
London SE26 5BN, UK
http://www.timberlake.co.uk

842 Greenwich Lane
Union, NJ 07083-7905, U.S.A.
http://www.timberlake-consultancy.com

This book was set in Palatino with the LaTeX programming language by the author.

Contents

List of Figures

Preface

This short introduction to macro-econometrics focuses on the concepts, tools and techniques needed to model aggregate economic data, here unemployment, wages, prices and money in the UK over a long historical period of approximately 150 years, usually 1860–2011, but somewhat shorter for a few of the time series. The basic framework draws on Hendry and Nielsen (2007), Hendry and Nielsen (2010), and emphasizes the need for general models to account for the complexities of economies and the magnitudes of the many changes that have occurred. That combination poses a major challenge for teaching elementary methods. Fortunately, despite the difficulties inherent in the distributional issues of estimators for non-stationary cointegrated time series with multiple location shifts and non-linear relationships, many of the key concepts can be explained using simple examples, then automated computer software can conduct the more complicated empirical modelling studies. The book is the outcome of a six-hour lecture course on Quantitative Economics delivered to second-year Philosophy, Politics, and Economics (PPE) undergraduates at Oxford University. A minimal background in elementary statistics is assumed, but developed quite rapidly while trying to minimize mathematical derivations.

Chapters commence with some 'Guide posts' as to their aims, and conclude with 'Key points' made during the discussion. Then there are four 'How to do it' Tasks at the end of every chapter, the more difficult of which are marked by an asterisk. The Tasks use *OxMetrics* (see Doornik, 2013b) and *PcGive* (Doornik and Hendry, 2013) as that is the only software that implements all the tools and techniques needed in the book. The more advanced simulations also require *Ox* (see Doornik, 2013a). The relevant subset of the data is available as QEHistData.xls, on the book's website, and the complete historical data set is UKHist2013.xls. Chapters also offer a small selection of exercises for readers to undertake to check their progress, or instructors to set, where answers are not usually provided.

The book is not about macroeconomics. As the assiduous reader will discover, much of the huge variation over long time periods in many aggregate variables does not fall under the purview of economic analysis, but is due to extraneous forces such as wars, changes in legislation, shifts in social mores, and technological, medical and financial innovations, which in turn are only partly affected by economics. The current vogue for seeking micro-foundations for such variables in terms of a 'representative' agent who is simultaneously, employed, unemployed, growing up and retired, rich and poor, etc. sits uneasily with the historical evidence.

The backgound research was originally supported in part by grants from the Economic and Social Research Council, and more recently by the Open Society Foundations and the Oxford Martin School. I am grateful to them all for the essential funding they provided, and to Jennifer L. Castle, Jurgen A. Doornik, Neil R. Ericsson, Vivien L. Hendry, Oleg I. Kitov, Grayham E. Mizon, John N.J. Muellbauer, Bent Nielsen, Felix Pretis and Tim Willems for many helpful discussions and suggestions, as well as to Charles Bean, Gavin Cameron, Christopher Gilbert, Fatemeh Shadman-Metha and John Muellbauer for various data series described in the Appendix.

The book was prepared in *OxEdit* and typeset in LaTeX using *MikTex*. Graphical illustrations, numerical computations and Monte Carlo experiments were done using *OxMetrics*, *PcGive* and *Ox*. The present release is *OxMetrics* 7.00 (July 2013).

Chapter 1
Macroeconomic data: evolution with abrupt change

Chapter 1 guide posts

1. Macro-econometrics concerns the analysis of data on whole economies, such as inflation, unemployment and gross domestic product (GDP).
2. This chapter uses a substantial annual data base for the UK starting in 1860, developed by economic historians: the role of graphs in describing such data is an important feature.
3. The last 150 years have witnessed huge changes in almost all aspects of life, especially living standards from technological, legal, medical and financial innovations, with consequential social and demographic shifts: Section 1.1.
4. The first focus is on wages and prices, which have risen roughly 700-fold and 100-fold respectively: Section 1.2.
5. A key feature of macroeconomic data is evolution interrupted by abrupt changes: Sections 1.3 and 1.4.
6. Overall, the purchasing power of wages (called real wages) has increased about 7-fold, approximately the same as average productivity per worker: Section 1.5.
7. That finding broadly matches a neo-classical model of firms equating the marginal revenue per worker with their marginal costs, but systematic departures between the data and the theory remain: Sections 1.6–1.9.
8. Trends and sudden shifts are common, so simple models of trends are developed in Sections 1.10 and 1.11: Chapter 2 will investigate shifts, many of which coincide with the major historical events discussed in Section 1.4.

1.1 Introduction

Knowledge of the historical context is an essential background to understanding macroeconomic evidence, and especially to undertaking any empirical investigations thereof. Many major changes have occurred historically, some because of dramatic events like wars and major technological and financial innovations—described in Section 1.4—others more gradually, but all affecting empirical models. Consequently, we first return to 1860 to trace what has happened since then.

So imagine you are living in London during 1860—just over 150 years ago. Queen Victoria has been on the British Throne for 23 years, and has another 41 to go; Abraham Lincoln has just been elected the 16th President of the USA, with Civil War looming; and Charles Darwin's *On the Origin of Species* is newly published. Infectious diseases are rife, and John Snow has only recently identified polluted water as the source of cholera. When employed, you would work up to 65 hours a week for less than 20% of modern (constant price, or real) incomes; you would often be ill from diseases that were not understood and had no known cures; living up to 10 to a room with little or no sanitation and no running water; hungry much of the time, and when you were unemployed, you and your family would starve; your children would be ill-clad and barefoot, mostly uneducated, many dying at birth or infancy; and even having survived to adulthood, on average you would die in your mid-40s.

Worse still, that situation describes someone living in the richest city of one of the richest countries in the world at the time. Although a small island nation, the United Kingdom then produced about 20% of World output (now around 3.5%). Almost everywhere else, living standards were lower (see Allen, 2011, for an excellent short economic history of the world), so it is little wonder Charles Dickens (*Great Expectations*, 1860) and Emile Zola (*Germinal*, 1885) wrote the sort of novels we know them for today. Bowley (1937) provides a more quantitative appraisal, consistent with their views.

How did society get where it is today from there? What statistical tools are needed to study evidence generated by a world that has changed so dramatically? We will consider models and methods for investigating macroeconomic data, seeking to explain the available empirical evidence on UK unemployment, wages and prices. A brief description of the data series and their measurements is provided in the appendix, Table 9: Hendry (2001) provides a more detailed explanation and sources.

This Chapter concerns the basic properties of macroeconomic data, characterized by evolution with abrupt, often unanticipated, changes. Chapter 2 discusses methods for 'taming' trends and breaks, to facilitate the development of sustainable empirical relationships. Then Chapter 3 considers how to characterize dependence between variables, and over time. Chapter 4 turns to an analysis of the two key ingredients of economics and statistics that make the backbone of macroeconometrics, before Chapters 5–8 provide applied studies in modelling UK unemployment, then wages, money demand, and prices respectively. Chapter 9 concludes.

1.2 Some facts about UK wages and prices

To investigate the huge changes that occurred over 1860–2011 in the UK, we will consider the time series of the relevant variables. First, some salient features:

- Nominal wages (denoted W) rose more than 680-fold, so an average wage of £1 per week in 1860 becomes £680 by 2011: that is a 68,000% increase in just over 150 years.

- Prices (denoted P) rose more than 98-fold, so £1 per item in 1860 cost £98 in 2011.
- Consequently real wages (as measured by the ratio W/P), or the purchasing power of earnings, rose approximately 7-fold.
- Industrial composition, technology, transport, wealth and income distributions, laws, education, social security, pensions, demography, health care, longevity, morbidity, sanitation, roles of women, social mores, and housing tenure have all also changed vastly over that period.

Change is the norm, both evolution and sudden shifts—modelling change is central to macroeconomic data analyses. A conventional way to present data evidence in macroeconomics is by time-series graphs. These show time on the horizontal axis, here from 1860 to 2011, with the magnitudes of the plotted variables recorded at each point on the vertical axis. Both wages and prices are measured as indices, with bases of unity in 1860, the former rising to 680 and the latter to around 98.

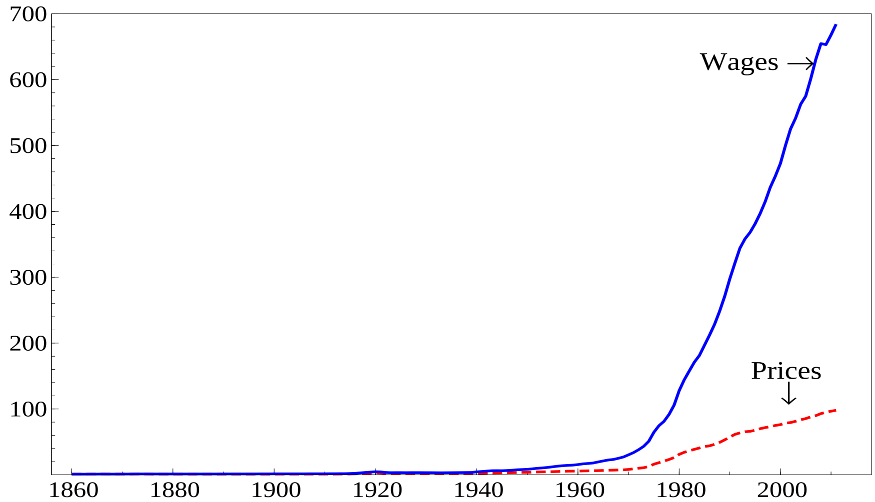


Figure 1.1
Wages and prices over 1860–2011.

Figure 1.1 shows the time series of nominal wages and prices over 1860–2011 (drawing this graph is set as an exercise in Task 1 Section 1.16). The early period looks relatively unchanging, but that is an artefact of later increases being so much larger because the graph shows absolute changes. Denoting a generic time-series process by $\{X_t\}, t = 1, \ldots, T$ where t is time for a period starting at 1 and ending at T, then most of the large increases in $\{W_t\}$ and $\{P_t\}$ came post-1975. Rising from 1 to 2 is 100% increase—but is dwarfed visually by a rise from 100 to 120, although that is just a 20% increase. The solution is to exploit the fact that $W_t,\ P_t$ must both be positive, so one can use logarithms of wages ($\log(W_t) = w_t$) and prices ($\log(P_t) = p_t$) to measure relative changes. Figure 1.2 shows these relative changes by plotting in logs, where the distance between successive 'tick' marks on the graph is a constant 20% change as shown (see Task 2 Section 1.17).

1.2.1 Properties of logarithmic transformations

A basic result in calculus is that changes in logs are relative to their level:

$$\frac{\partial \log X}{\partial X} = \frac{1}{X} \tag{1.1}$$

If $X = 1$, one unit increase in X to 2 is 1/1, which is 100%. When $X = 100$, 1 unit increase in X to 101 is 1/100, or 1%. For a 100% change, one needs to go from $X = 100$ to $X = 200$. However, equation (1.1) only holds for small changes: $\log 1 = 0$ but increasing to $\log 1.5 = 0.4055$ is just over 40% not 50%. Shops record a change from, say, £1 to £1.5 as a 50% increase, whereas going from £1.5 to £1 is a 33% reduction, so such changes are not symmetric (this is why the RPI is being abandoned). Logs show symmetric changes—going from $\log 1.5$ to $\log 1$ is a 40% fall. We will use capitals like X for the variables in their original units, and lowercase letters like x for their logs, as used above for wages and prices.

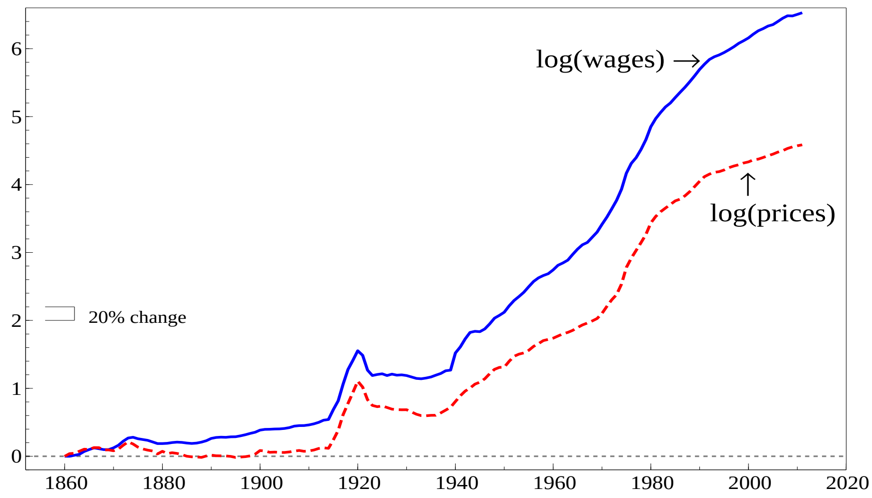


Figure 1.2
Log wages and log prices over 1860–2011.

1.3 Evidence from data graphs

The graph of the logs of the variables in Figure 1.2 now reveals that the early period (before the First World War) was in fact far from unchanging, with prices first rising then falling throughout the remainder of the 19th Century. Moreover, wages and prices were also very volatile between 1914 and 1945, initially both rising dramatically, followed by large falls then a gradual downwards drift in the interwar period. Overall, wages have grown much faster than prices, but neither shows constant growth over the 150 years.

To describe the average growth of wages and prices, we can fit a trend line as shown in Figure 1.3 for w_t. This line corresponds to a simple 'model' of the form:

$$x_t = \alpha + \beta t + e_t \tag{1.2}$$

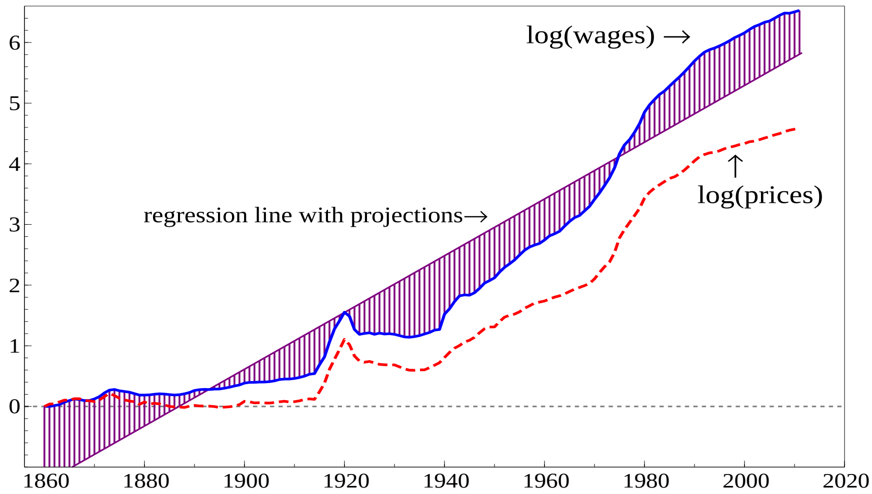


Figure 1.3
Log wages and log prices with one trend line.

were α is the intercept (the average value of x_t when $t = 0$), and $\beta = \partial x_t/\partial t$ is an assumed constant rate of growth. The unexplained component is denoted by e_t, but we will not ascribe any statistical properties to e_t yet. Fitting a simple trend line as in (1.2) can be done by finding the 'best fit', namely minimizing the sum of squared values of e_t (i.e., $\sum_{t=1}^{T} e_t^2$). We first fit that trend line to w_t using *OxMetrics*, showing the deviations $\widehat{e}_t = x_t - \widehat{\alpha} - \widehat{\beta}t$ as projections from the line, where $\widehat{\alpha}, \widehat{\beta}$ are the values that minimize $\sum_{t=1}^{T} \widehat{e}_t^2$. The squares of the projections shown were minimized in the sense that moving the fitted line up or down, or altering its slope in either direction, will produce a larger sum of squared projections. It would be hard to check that claim by hand, but trivial on a computer.

It is fairly obvious from Figure 1.3 that a single trend line does not describe the evidence on w_t at all well: there are large and systematic deviations from the fitted trend, which indeed crosses the data line only three times in 150 years. A similar result holds for a single trend fitted to p_t. For better descriptions, we can try dividing the overall sample into sub-periods, and find the fitted trends within each. Here we will use six periods of about 25 years each, and fit separate trend lines within each 'epoch'. Doing so using *OxMetrics* delivers Figure 1.4.

Each sub-period fit is naturally much better, but overall reveals a different problem: the trend lines do not all have the same slopes. Rather, growth rates of wages and prices have changed substantially over the period.

Lines selected by 'best fit' can be applied to any data, so let me act like Picasso and autograph my pretty picture, then fit a trend line to my signature as shown in Figure 1.4 (see Task 2 Section 1.17). Carrying out the same operation of fitting six trend lines to prices, yields similar, but not identical, patterns of slopes as also shown in Figure 1.4. Changes in slopes are again apparent, and even with six separate trend lines, several periods are not well described (especially 1910–1934, where at least two more slope changes are needed). Such findings raise four questions.

Why were there so many changes in the growth rates?
How could one know that 6 regressions, rather than 1, or 12, were required?

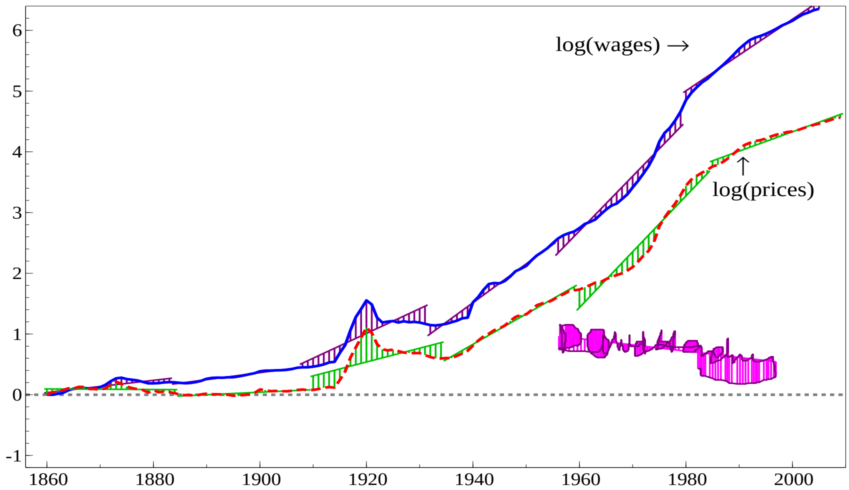


Figure 1.4
Log wages and log prices each with six trend lines.

How can such changes be modelled?
What else is needed to explain such changes?
Section 1.4 tries to answer the first by considering the many major historical events that affected the UK over 1860–2011. However, answers to the other three questions will take much of the rest of the book.

1.4 Major historical events affecting the UK over 1860–2011

We characterize these major historical events by five sub-groups:
[A] dramatic shifts;
[B] key financial innovations;
[C] important societal changes;
[D] technology advances; and
[E] policy regime shifts.

[A] Dramatic shifts that occurred over 1860–2011 include:
World War I (WWI); the 1918–21 flu' epidemic; the 1919–21 crash; the 1926 general strike; the 1930's Great Depression; World War II (WWII); the 1970's oil crises; and of course, the 2008–2013 financial crisis and 'Great Recession'. These major events often altered previous relationships between variables, including between wages and prices.

[B] Key financial innovations and changes in credit rationing took place:
personal cheques (introduced in the 1810s); telegraph (economizing on multiple bank accounts during the 1850s, as discussed by Alfred Marshall, 1926); credit cards (1950s) and ATMs (1960s); as well as deregulation of the banking sector (from the 1980s). These affected the volumes of 'money' in circulation as well as greatly widening access to finance.

[C] Many important societal changes occurred, some noted above, including: demography, family size and structure; laws, social mores, roles of women; health care and sanitation, resulting in greater longevity and lower morbidity; education, social security, pensions, housing tenure; wealth and income distributions.

[D] Huge technology advances were implemented:
industrial composition was radically changed as the industrial revolution staples of cotton cloth, coal mining, steel making, and shipbuilding almost all vanished in the UK, to be replaced by electricity, gas and nuclear power; computers, mobiles, GPS, internet and many forms of communications; major medical developments like vaccination, antibiotics, biotechnology and DNA; new forms of transport with both cars and planes, after the boom in, then demise of, canals and later railways.

[E] Many policy regime shifts took place:
from the gold standard till Bretton Woods (1945, with some short departures during the Great Depression and World Wars); then floating exchange rates (till about 1973); followed by a succession of Keynesian, Monetarist, then inflation-targeting policies; the creation of the European Union, then the Euro; and on to Quantitative Easing; to list a few of the most salient.

With such a catalogue of major events, it is not surprising that many economic relationships shifted. Empirical modelling must address all these large intermittent shifts, many of which may have been unanticipated by even the most sophisticated economic actors of the time. To emphasize the pervasive and pernicious occurrence of major shifts, Barro (2009) estimates very high costs from what he denotes 'consumption disasters'. He finds 84 events over approximately the last 150 years with falls of more than 10% per capita, cumulating in a total duration of almost 300 'bad' years across his sample of 21 countries, mainly due to wars, with a marked reduction in the frequency of 'bad' years after World War II.

1.5 Behaviour of real wages

We next examine what happened to real wages, which are essentially a synthetic time series to represent wages measured at constant prices, calculated here as $\log(W_t/P_t) = (w_t - p_t)$, and shown in Figure 1.5. By definition, $w_t - p_t$ is the vertical difference between the two series shown in Figure 1.2.

The graph shows that real wages have experienced substantial growth: a move between each small tick mark is about 5%, cumulating to about a 7-fold increase over the period. Such a graph shows that we are much better off today than in 1860, though it does not explain why real wages have risen so much. Moreover, many crucial aspects of modern life, including the huge improvements in sanitation, medical care, reductions in the killer diseases and increased longevity, are additional to improvements in real wages, an aspect correctly stressed by Crafts (2002).

The large impacts of the two world wars are clearly visible in terms of a sharp rise on commencement (1914, 1939) then a fall back on termination (1918, 1945). The reductions in real wages due to the 1970's Oil Crises and the ensuing high inflation and social unrest are also visible (a higher price for a key imported commodity acts like a tax). More generally, the growth rate is not constant, and increased markedly after World War II. As figure 1.6 shows, a single trend line does not describe the history well: there are systematic deviations between the line and

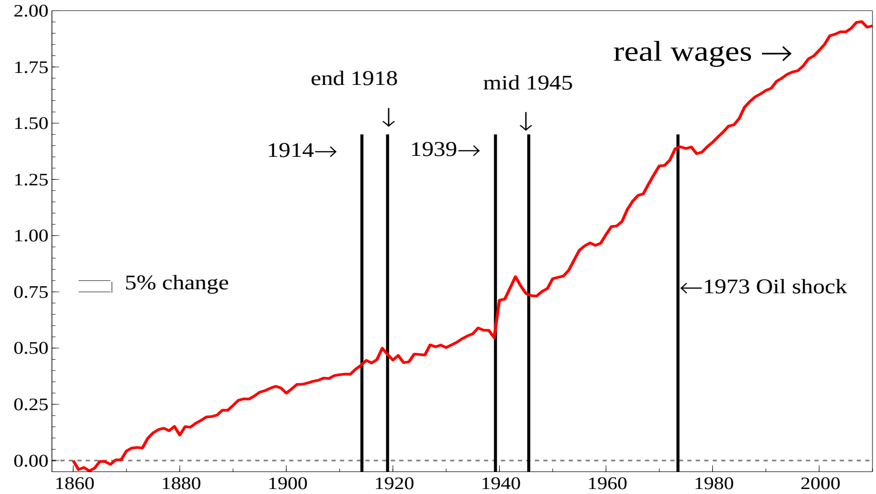


Figure 1.5
Log real wages, $w - p$, over 1860–2011.

$(w_t - p_t)$. Indeed, such deviations only change sign twice in 150 years, so are very far from random. As an aside, notice that the data lie above the trend line both when growth is *slow* at the start and when it is *fast* at the end: so called 'gaps' between variables and trend lines are not interpretable when growth rates change.

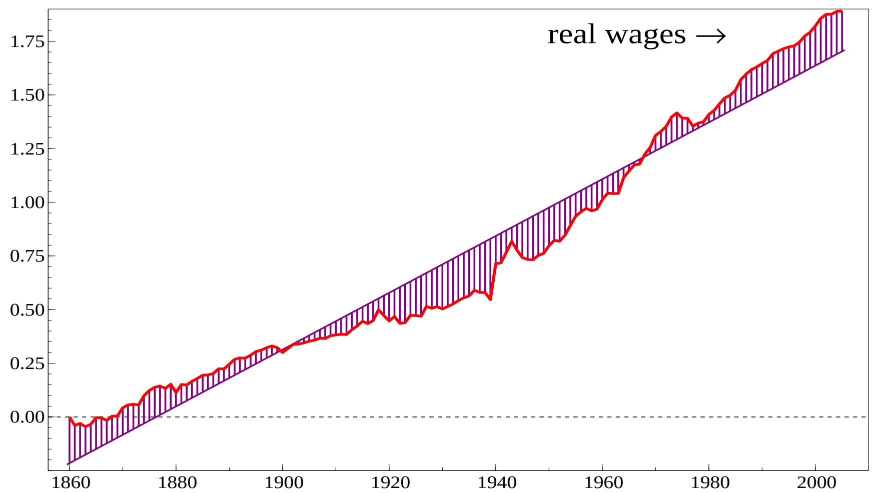


Figure 1.6
Log real wages with one trend line.

There are four distinct sub-periods or 'epochs', each of roughly 40 years, corresponding to changes in the growth rate: moderate growth till about 1900, then slow till the start of World War II, somewhat faster till the late 1960s in the post-war reconstruction, then moderating again till the end of the sample. That suggests fitting four separate trends: using *OxMetrics*, the outcome is shown in Figure 1.7, and provides a far better description, although systematic departures remain as can be seen.

Nevertheless, why have real wages grown at all, and why at different rates in different time periods? To address these issues, we digress into some economic theories of wage determination to guide which variables might contribute to such an explanation.

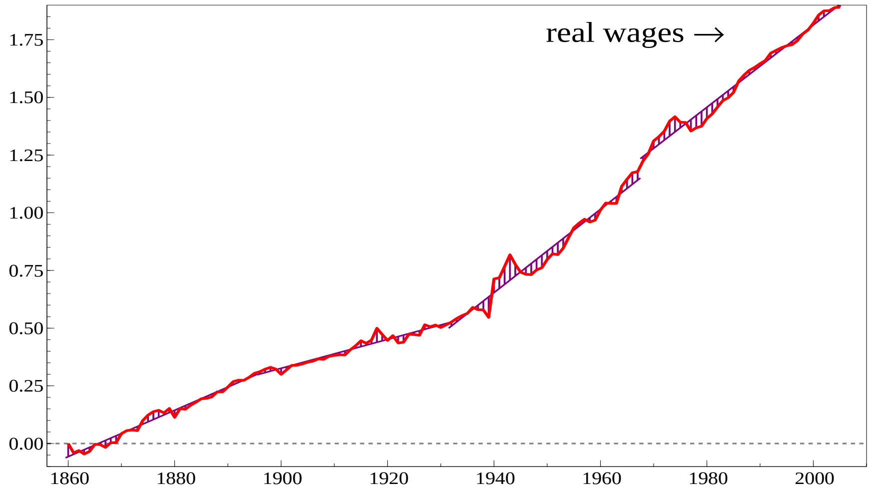


Figure 1.7
Log real wages with 4 regressions.

1.6 Theories of wage determination

There are many theories of how wages are determined, from Malthusian subsistence (any increase in real wages is offset by an increased population driving wages back to subsistence); Walrasian labour-market clearing (wages are set at the level where the demand for labour equals the supply); Keynesian nominal rigidities (social norms and worker pressure act against wage and price cuts as those increase the real value of debts owed by the poor); micro-founded search and job matching with important roles for imperfect information and adverse selection (employers do not know the true characteristics of prospective workers, but workers know themselves); to efficiency wage theories (it is advantageous to pay above the market clearing wage if that improves the health and hence the productivity of the workers), among others.

All of these may well be part of the explanation in various times and places, although not all are crystal clear about whether their relevance is to nominal or real wages. However, taking them in turn for our period and country, the first has so far failed to happen, partly from the historical 'accidents' of the Industrial Revolution and ensuing technological progress greatly increasing the output per person of physical goods, and partly the development of the 'New World' as a 'bread basket' (see Allen, 2009, for an explanation of the Industrial Revolution, and Allen, 2011, for a global economic history). We consider the second theory at greater length below. The third is at best part of the story: the dramatic falls in wages and prices evident in Figure 1.2 must have been known to John Maynard Keynes at the time, and hence pre-date Keynes (1936). The last two are primarily useful in explaining microeconomic phenomena.

By this process of elimination, we are left with the so-called 'neo-classical' theory of wage determination: profit-maximizing competitive firms hire workers till the real marginal cost of labour (each new employee's additional total real wage cost, MC) equilibrates to their real marginal revenue product (MR), so that MC=MR. Let G denote aggregate real gross domestic product (GDP), and L denote total employment, then WL is the nominal wage bill and PG is nominal GDP, so the theory entails:

$$\frac{\partial(WL/P)}{\partial L} = \frac{\partial(PG/P)}{\partial L} \tag{1.3}$$

We will consider the determinants of MR then of MC in turn, commencing with output determination in Section 1.7, costs in Section 1.8 then linking real wages and productivity in Section 1.9.

1.7 Production relationships

Production relations linking outputs of goods and services to inputs of resources are a primary determinant of the demand for labour, although so is aggregate demand as there is no point in producing goods that cannot be sold. Economists know remarkably little about the microeconomic detail of such relationships outside of agriculture. Indeed, with multi-product firms and multi-tasking workers it is far from obvious how to attribute either costs or value added to individual outputs or inputs. However, at the aggregate level, constant returns to scale seem likely (see e.g., Houthakker, 1956) so an $x\%$ increase in all aggregate inputs (if such was feasible) would lead to an $x\%$ increase in all outputs. When constant returns to scale and no technical progress both hold, marginal values are proportional to averages, MR $\propto$ AR, but the absence of technical progress is implausible.

For simplicity, we assume the production function has the Cobb–Douglas form shown in (1.4): see Cobb and Douglas (1928). Outputs, G_t, as measured by value added are proportional to the product of the inputs of capital, denoted K_t, and labour, L_t, so zero net output is produced unless both inputs are positive. The amount produced increases over time by $A(t) = \exp(at)$, which is called 'disembodied' technical progress, assumed to occur at the constant rate a per annum, noting that most technical progress is actually embodied in capital and labour:

$$G_t = \exp(at)L_t^{\lambda}K_t^{(1-\lambda)} \quad 0 < \lambda < 1 \tag{1.4}$$

In (1.4), λ will turn out to be the share of labour in GDP, so $1 - \lambda$ is the share of capital. K_t should be measured by the flow of 'quality adjusted' capital inputs taking account of scrapping and the differing efficiencies of different technology vintages, but such data are not available. Equally, L_t should be measured by the flow of 'human capital' hours of input; but again unfortunately, we have very limited data on hours worked, although we do know they fell greatly over our period, or paid holidays, which rose considerably, and even less on 'human capital' itself (employment adjusted for embodied skills and knowledge), which rose considerably with improved knowledge and increased education. Thus, K_t is measured here by the total capital stock in the UK (calculated by cumulating gross investment and assuming a rate of scrapping and obsolescence), and L_t by total full-time employment. Thus, essentially all forms of technical progress, improved knowledge and increased education must be 'picked up' by the catch-all $A(t)$, so any interpretation must be as 'not-explained-elsewhere' (i.e., representing our ignorance: compare Solow, 1956).

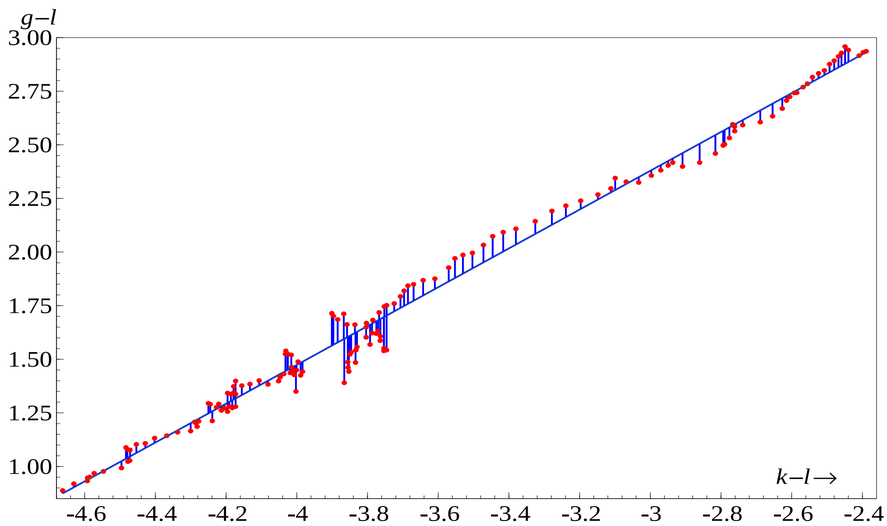


Figure 1.8
'Production function' over 1860–2011.

Taking logs in (1.4):

$$g_t = at + \lambda l_t + (1 - \lambda)k_t \tag{1.5}$$

and subtracting l_t from both sides then rearranging:

$$(g - l)_t = at + (1 - \lambda)(k - l)_t \tag{1.6}$$

where $(g - l)_t$ approximates real average revenue product per worker (AR). Figure 1.8 records the resulting 'production function' linking $(g - l)_t$ and $(k - l)_t$ represented by (1.6), with deviations of $(g - l)_t$ relative to the regression line shown (see Task 3 Section 1.18). The outcome is quite close to linear apart from 1920–1940, and consistent with constant returns to scale but slow disembodied technical progress. Although fitting (1.6) requires the use of methods not introduced till Chapter 5.5 below, doing so delivers estimates of $\widehat{a} \approx 1$%p.a., and $\widehat{\lambda} \approx 0.23$, so capital earns a share of roughly a quarter of GDP.

Figure 1.9 shows the 3-dimensional relationship over time between $(g - l)_t$ and $(k - l)_t$, and reveals different rates of technical progress pre-1918 and post-1945 (which can be estimated at roughly 1%p.a. before and 1.7%p.a. after), as well as emphasizing the large distortions in the interwar period, but not showing any long-run departure overall from constant returns to scale represented by the straight line.

To conclude this section, assuming technical progress is not affected by changes in employment, from (1.5) the marginal revenue on the right-hand side of the log

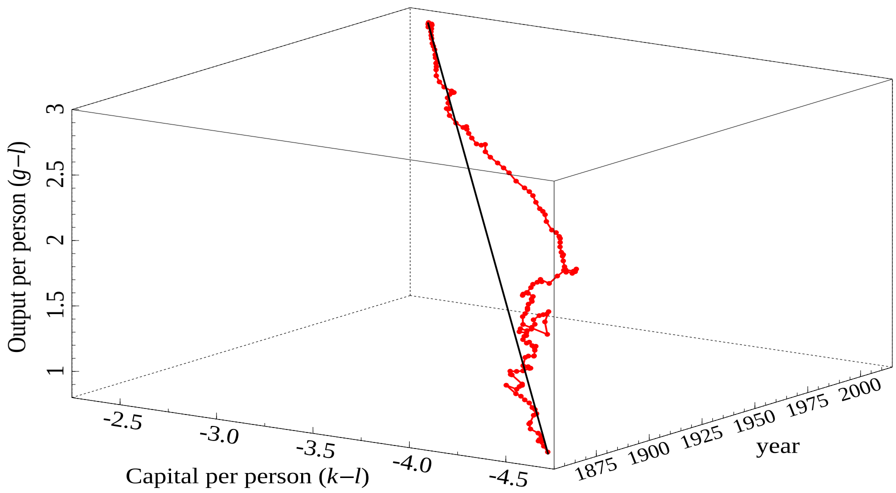


Figure 1.9
3-dimensional 'production function' over 1860–2011.

version of (1.3), becomes:

$$\frac{\partial g_t}{\partial l_t} = \lambda + (1-\lambda)\frac{\partial k_t}{\partial l_t}.$$

This implies a marginal revenue product of λ plus $(1-\lambda)$ times the change in capital per worker, which is also the average revenue product, $(g-l)_t$ when $a = 0$ and $\partial k_t/\partial l_t = (k-l)_t$. However, it is possible that changes in employment do affect technical progress (e.g., labour saving inventions), and may also change λ, issues to be addressed empirically.

1.8 Cost relationships

We now turn to the left-hand side of equation (1.3). It is often assumed that:

$$\frac{\partial(WL/P)}{\partial L} = W/P \tag{1.7}$$

but that derivation requires that changes in L do not affect W/P, so firms are 'price takers' in the labour market. Such an assumption seems most unlikely in the macroeconomy, and in a number of theories, some of which we address below, unemployment affects W/P.

Direct average costs per worker, AC, comprise W, labour taxes τ and benefits b (such as pensions), but we have no time-series data on the last of those, and little on τ till relatively recently. Figure 1.10 records the available data on some auxiliary time series of potential relevance, comprising days lost through strikes (s_t), real National Insurance Contributions ($nic_t - p_t$), membership of Trades Unions (tu_t), and unemployment benefit payments as 'replacement ratios' relative to average earnings ($repl_t$), all in logs, with their changes shown in the second row of graphs.

As a measure of employment taxes, $nic_t - p_t$ is now a substantial component of τ. Unfortunately, many of these time series are only available over relatively short

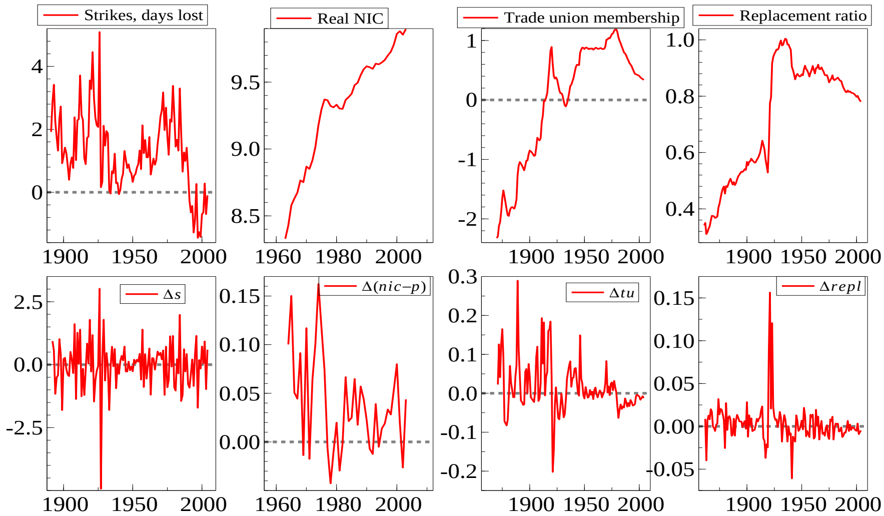


Figure 1.10
Log levels and changes in: days lost through strikes; National Insurance Contributions (NICs); Trades Unions membership (TU); and replacement ratios.

historical time spans. In logs, allowing for taxes and benefits, the right-hand side of (1.7) becomes $w_t - p_t + \tau_t + b_t$, so the long-run constant-price relationship for $AC = AR$ is:

$$w_t - p_t = g_t - l_t - (\tau_t + b_t) \qquad (1.8)$$

Then, $(w_t - p_t - g_t + l_t)$ is the wage share, or real unit labour costs, which (1.8) suggests should only change with $(\tau_t + b_t)$. If so, Trades Unions' membership, strikes, and unemployment benefits as shown in Figure 1.10, as well as employment legislation etc., must mainly impact on the unemployment rate, $U_{r,t}$, rather than on $(w_t - p_t)$–which is a testable hypothesis.

1.9 The relationship of real wages to average productivity

In (1.8), $(w_t - p_t)$ is the real product wage, measured relative to GDP prices rather than (say) consumer prices. Figure 1.11 plots log real wages $(w-p)_t$ and log average productivity $(g - l)_t$ over time. Matching the analysis in sections 1.7 and 1.8, the two variables show very similar long-run trends, and even match fairly well over the four epochs identified in section 1.5. Most of the rise in real wages has been due to increased output per worker. Of course that pushes the explanation back one step—we now need to explain increases in $(g - l)_t$, which seem to be due to increases in $(k-l)_t$, which cumulates past net investment.... We will not have space to push back explanations to 'fundamentals' in this book, but needing to extend a system being analyzed to explain 'explanatory variables' is a standard econometric problem. Moreover, such extensions have to be consistent within the system as a model of the macroeconomy.

Although the long-run relationship is close, there are persistent deviations between the lines as $(w - p)_t$ is below $(g - l)_t$ over the entire period 1865–1920 and again over 1925–1940, then remains above for 1940–1980, but ends by being close

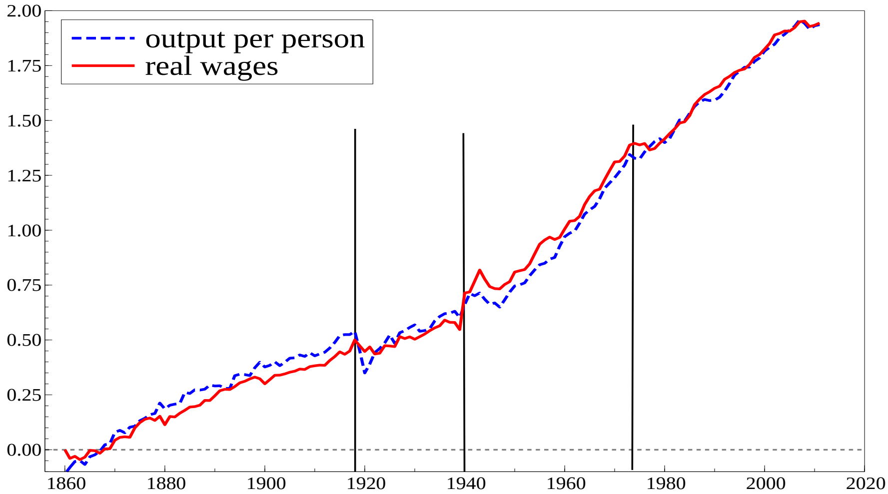


Figure 1.11
Log real wages and log average productivity over 1860–2011.

from then onwards. This is another standard econometric problem—needing an extension to explain systematic mis-matches. We will return to that issue after considering a third econometric problem—how to handle trends.

1.10 Deterministic trends

Most data graphs shown so far have revealed trends, so we need tools for handling many types of trends, some of which are deterministic, and some stochastic as we now explain: we will return below to the added problems that the 'trend is your friend till it doth bend'.

We begin with constant deterministic linear trends defined by:

$$x_t = \alpha + \beta t \quad \text{for} \quad t = 1, 2, 3, \ldots, \tag{1.9}$$

Then x_t grows at the rate β, because deriving the outcome one period earlier:

$$\begin{aligned} x_t &= \alpha + \beta t \\ &= \big(\alpha + \beta(t-1)\big) + \beta \\ &= x_{t-1} + \beta \end{aligned} \tag{1.10}$$

where x_{t-1} is the previous value, or lag, of x_t. Subtracting x_{t-1} from both sides of (1.10):

$$x_t - x_{t-1} = \Delta x_t = \beta \tag{1.11}$$

where Δx_t is the change in x_t. For a process like (1.9), the difference Δx_t is constant at β, and is trend free. When x is the log of a level, Δx_t is the growth rate.

As a possible example, Figure 1.12 plots the changes in average productivity. However, these do not look very constant, consistent with (1.6), which entails:

$$\Delta(g - l)_t = a + (1 - \lambda)\Delta(k - l)_t$$

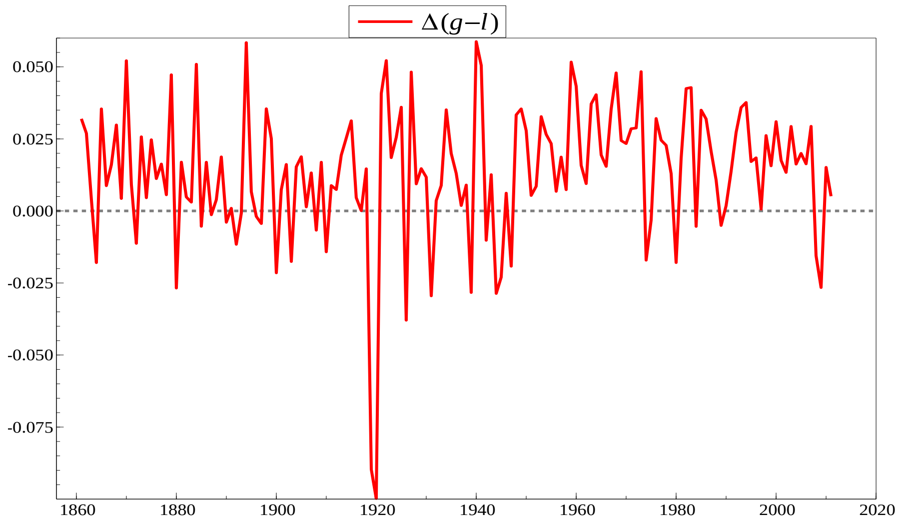


Figure 1.12
Changes in average productivity $\Delta(g-l)_t$ over 1860–2011.

so $\Delta(g-l)_t$ varies with $\Delta(k-l)_t$, which is net investment per employee, and is bound to change with technology and over business cycles, as well as with special events such as wars and major policy changes. Consequently, we now consider what are called 'stochastic trends'.

1.11 Stochastic trends

Returning to equation (1.10), re-write it with an added error, or 'shock', as:

$$x_t = x_{t-1} + \beta + \epsilon_t \tag{1.12}$$

In (1.12), ϵ_t is the random shock, which we assume to be:

$$\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1.13}$$

denoting an **I**ndependently distributed **N**ormal random variable with a population mean, or expectation, of zero, so $\mathsf{E}[\epsilon_t] = 0$, and a variance of $\mathsf{E}[\epsilon_t^2] = \sigma_\epsilon^2$, as illustrated in Figure 1.13.

From (1.12), $\Delta x_t = \beta + \epsilon_t$, so the change Δx_t now varies randomly around β. $\{x_t\}$ is called a random walk, with drift at the rate β: equity prices are often treated as random walks with $\beta = 0$ because the change is then unpredictable, precluding a free lunch. However, allowing for an error on the change has some crucial implications, which we now explore.

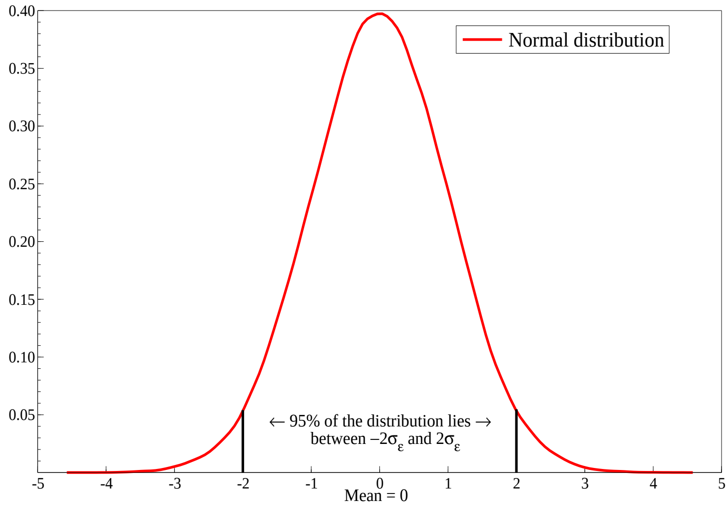


Figure 1.13
Normal distribution with mean $\mathsf{E}[\epsilon_t] = 0$ and variance $\mathsf{E}[\epsilon_t^2] = \sigma_\epsilon^2 = 1$.

1.11.1 Properties of stochastic trends

At each point in time, x_t has cumulated all the past ϵs because:

$$\begin{aligned}
x_t &= x_{t-1} + \beta + \epsilon_t && \text{[by definition]} \\
&= \left(x_{t-2} + \beta + \epsilon_{t-1}\right) + \beta + \epsilon_t && \text{[one period earlier]} \\
&= x_{t-2} + 2\beta + \epsilon_t + \epsilon_{t-1} && \text{[rearranging]} \\
&= \vdots && \text{[repeatedly lagging]} \\
&= x_0 + t\beta + \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1 && \text{[at time zero]} \\
&= \alpha + \beta t + \textstyle\sum_{r=1}^{t} \epsilon_r && \text{[changing notation]} \\
&= \alpha + \beta t + u_t && \text{[looking like (1.9)]} \qquad (1.14)
\end{aligned}$$

where $\alpha = x_0$ and $u_t = \sum_{r=1}^{t} \epsilon_r$. Then, (1.14) shows that x_t cumulates all past errors, $\sum_{r=1}^{t} \epsilon_r = u_t$, so will 'wander widely', as well as trending at the same rate β as (1.9). The cumulative errors will play an important role in later chapters.

1.11.2 The concepts of stationary and non-stationary processes

A time series is stationary if the distribution from which the data are drawn remains the same over time. In particular, the mean and variance of that distribution need to remain constant. If the distribution or any of its moments change over time, then the process is non-stationary.

The graphs of time series seen so far are all distinctly non-stationary, in that their means have not remained constant, but have usually trended over time. Figure 1.14 records the histograms and approximating densities (dashed and dotted

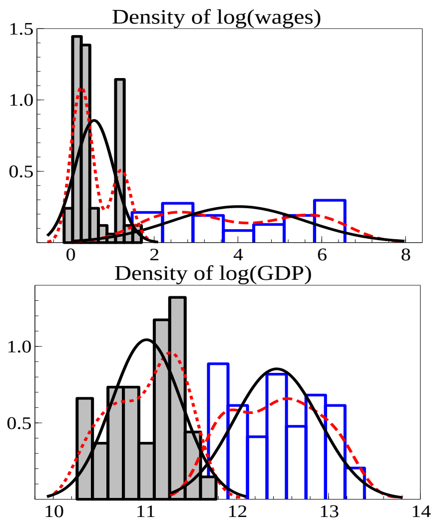


Figure 1.14
Histograms and densities of w_t, p_t, g_t and l_t, shaded pre-WWII.

lines) of four of the variables, w_t, p_t, g_t and l_t, split before and after 1940.[1] The former densities are shaded, the latter are not. There are large shifts in the distributions, with both the means and the variances changing between the two periods, as well as revealing apparent non-normality, where the closest normal distributions are shown as solid lines (see Task 4 Section 1.19).

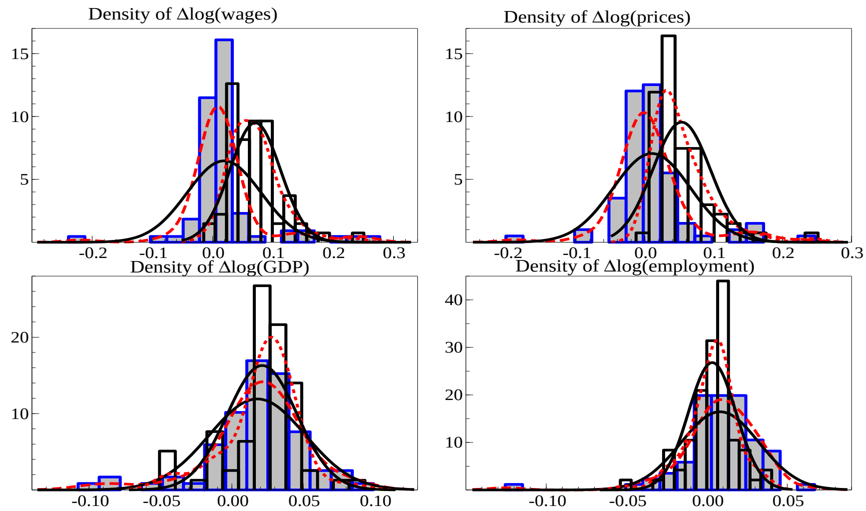


Figure 1.15
Histograms and densities of Δw_t, Δp_t, Δg_t and Δl_t, shaded pre-WWII.

[1] Histograms allocate observations to each interval, so depend on how those intervals are chosen: the approximating densities smooth across the intervals using a weighted average where the weights are largest for observations near the centre of the interval and converge on zero for observations in distant intervals.

By way of contrast, Figure 1.15 records the histograms and approximating densities (dashed and dotted lines) of the first differences of the 4 variables, Δw_t, Δp_t, Δg_t and Δl_t, which still change between the two periods, but by much smaller amounts and are also closer to normality within each period. Nevertheless, it cannot be assumed that differences are necessarily stationary: trends can change, or drifts can shift.

1.12 Some additional background

Although variables such as p and w have changed vastly over our sample, when 'scaled' by the magnitude of the economy of their time, other variables have fluctuated considerably but still have magnitudes recognizable today. Four of these are the amount of broad money in circulation, measured by M4 as in Ericsson, Hendry, and Prestwich (1998) and shown relative to nominal GDP (denoted by $m-g-p$ in logs) which we will analyze in Chapter 7; the ratio of nominal national debt to nominal GDP (denoted by $N/(PG)$, not in logs); the long- and short-term interest rates (R_L and R_S respectively); and the purchasing power of £ sterling against an index of world currencies (weighted by UK export shares, and denoted ppp, in logs). These are shown in Figure 1.16.

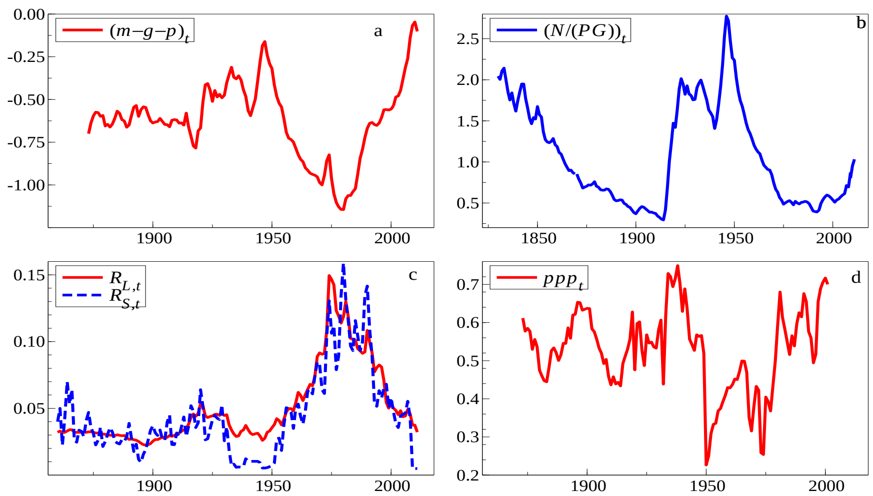


Figure 1.16
Time series of $(m-g-p)_t$, $(N/(PG))_t$, $R_{L,t}$ and $R_{S,t}$, and ppp_t.

As can be seen in panel a, the relative amount of money in circulation has moved over a wide range, rising by 60% to a peak just after WWII, before more than halving to a trough in 1980, then recovering to well above its previous peak. Chapter 7 investigates the demand for money.

Panel b shows a longer time series for $(N/(PG))$ to emphasize the non-economic aspects of its behaviour. Starting at a high level after the Napoleonic Wars, which has a similar value to the ratio in 1918, it declined slowly till 1860, reaching a low in 1913 before climbing back to a peak in 1946, declining to roughly the same level as pre-WWI in 1990, then rising again after that downturn and the 2008 recession.

Almost all the variation is due to wartime expenditures, peacetime reductions induced by rising nominal GDP, and post-WWII fiscal stabilizers. The present ratio is well below the average of the last 200 years, when higher levels did not prevent the UK from funding the Industrial Revolution and becoming a major world power.

Panel *c* reveals that R_L is like an average of R_S pre-WWI, diverges during the interwar and immediate postwar years (possibly Keynes's 'liquidity trap'), before both leap with the inflations of the 1970s and 1980s, then gradually decline into the 21st Century, again diverging with the 'Great Recession'.

Finally, the purchasing power of £ sterling has seen highs and lows, but is not far from its original value. Thus, three of the four variables have ended with similar values to their initial magnitudes: the fourth ($N/(PG)$) would have done so if the first date shown had been 1860. We will develop concepts to help explain such 'relatively stable' behaviour against the background of huge changes in the economy and society, where the levels of M and N have increased by almost 200,000% and 170,000% respectively.

1.13 The way forward

This Chapter looked at the basic properties of macroeconomic data. We found that variables like aggregate wages and prices are characterized by evolution with sudden shifts. Many levels have increased dramatically over the last 150 years, and their shifts often coincided with major historical events. Thus we need tools for analyzing models of data that have changed so greatly. Chapter 2 will discuss models of trends and breaks and methods for 'taming' them in order to facilitate the development of sustainable empirical relationships. Then Chapter 3 will consider how to characterize the high levels of dependence between macroeconomic variables, and also over time. Chapter 4 will briefly review the two key ingredients of economic theory and statistical analysis that form the backbone of macroeconometrics. The heart of the book comprises the four empirical modelling studies for UK unemployment in Chapter 5, real wages in Chapter 6, money demand in Chapter 7, and price inflation in Chapter 8.

1.14 Chapter 1 key points

(A) There have been huge changes in nominal wages and prices over the last 150 years in the UK (and most other countries, though at different times): approximately 700-fold and 98-fold respectively.
(B) Real wages have increased greatly (about 7-fold), and that rise has been close to the rise in average productivity per worker.
(C) Such a finding roughly matches the simplest neo-classical model of equating marginal revenue per worker with marginal costs, subject to a number of assumptions such as constant returns to scale.
(D) Nevertheless implementing that theory leads to systematic departures from the available data, which admittedly omit important changes in hours, holidays, education, labour force composition and skills, as well as labour taxes and benefits.
(E) Many dramatic historical events, technological, legal and financial innovations,

and social and demographic developments occurred over the period.
(F) The measured growth rates of many variables have changed on several occasions, leading to shifts in their distributions.
(G) In general, macroeconomic data manifest both evolution and abrupt changes: trends and sudden shifts are common, making most time series non-stationary.
(H) We formulated simple models of trends, both deterministic and stochastic.
(I) Other variables like broad money and the national debt have also increased dramatically, but have reverted to similar ratios to GDP.

The next chapter will describe ways of 'taming' such trends, then analyze and also model sudden shifts—another manifest feature of Figure 1.16.

1.15 Introduction to the Tasks

The 'How to do it' Tasks are mainly based on *OxMetrics* (see Doornik, 2013b) and *PcGive* (Doornik and Hendry, 2013) as that is the only software that implements all the tools and techniques needed in this book, but nevertheless provides an easy-to-use menu-driven approach to plotting graphs, transforming data, conducting regression calculations, and simulating statistical distributions.

These Tasks of replicating what the chapters discuss assume the use of a PC with a mouse, and focus on accessing operations by clicking on the icons on the third row from the top of the screen using a mouse, although access can usually also be achieved by holding down the Alt key and pressing the relevant **L**etter (e.g., Alt+**G** for the **G**raphics dialog), or by clicking on the relevant `Menu` button on the second row from the top of the screen. Buttons, menu and dialog choices are shown in `Typewriter` font. Throughout, 'click' denotes the left mouse button, unless the right button is specifically requested to carry out that part of the task; and on-screen selections are also denoted in typewriter font as shown for `Menu` just above. Variables in the database are named as 'W', 'Dwp' and so on. Readers familiar with *OxMetrics* should be able to conduct the Tasks as they proceed. The Tasks are placed after the 'Key Points' sections so they can be skipped if desired. However, many Tasks lead to developments in applying the econometric tools relevant to macroeconomic data, as well as explanations of tests and procedures that would distract from the main text. Conversely, the Tasks deliberately increase in difficulty as the book progresses, and the most difficult are denoted by *.

1.16 Task 1: Loading and graphing data

Once *OxMetrics* is installed, click on its icon to start: its help system can be accessed by clicking on the `Help` button on the second row from the top of its screen when the program is in focus: we will call that row the 'Command line' henceforth.

The first task is to load the data set. Right click with the mouse on the folder `Data` (under 'Documents' on the left-hand side of the screen), select `Open Data`, and click on 'UKHist2013.xls' to load. This data set should now be listed just below `Data`. Click on 'UKHist2013.xls' to bring it into focus.

The next task is to draw Figure 1.1 for nominal wages (W) and prices (P) over 1860–2011. Capitals (W) and lower-case (w) names are distinguished by *OxMetrics*. Click on the graphics icon (9th from the right on the third row from the top, henceforth called the 'Icon line'), or use Alt+**G**, or click on the '`Model`' button on the 'Command line' and then on `Graphics`. The dialog has three columns, labelled `Selection` on the left (presently empty as nothing is selected), double-direction arrows ('<<' and '>>') in the centre, and `Database` on the right, which lists all the available time series. To find a variable, click inside the `Database` column, then type the first letter of its name. Thus, type 'W' and it will be highlighted (if you type 'W' again, 'WPOP' will be highlighted). To add 'W' to the selection for graphing, either click on the centre column arrow '<<', which should be highlighted as well, or double click on the name highlighted. Now 'W' should appear in the `Selection` column. Repeat for 'P' so both are selected and now click on `Actual series` button on the bottom left of the dialog to see the graph. As they are indexes, the units of W and P are arbitrary, and have been set to unity in 1860.

To edit a graph, either double click on it, or click on the `Edit` button on the 'Command line' and then `Edit Graph`. That brings up the dialog where you can change names of variables, line choices, font sizes, etc. For example, double click on the blue line shown opposite `Style` for P and change it to a dashed or dotted line. When editing is complete, either press the 'Esc' key on the keyboard, or click on the red cross at the top right of the dialog to close that window. To type on the graph, literally just do that when it is in focus, shown by a jagged line round the figure. So, type (e.g.) 'Wages': after the first letter, a `Graphics Text` box will appear. Graphs in *OxMetrics* support the LaTeX typesetting language, so to add an arrow pointing to the wages line, type '$\rightarrow$' (which will show on the graph as →) and then click on `OK`. Once text shows on a graph, you can move it (click the text and hold the mouse button down, then shift that text to where you want it), or edit it (double click on it, and you will see it listed under `Text` on the `Edit Graph` dialog. Hopefully your graph looks like Figure 1.1. The `Blue bent arrow` on the Icon line can undo any mistakes.

Save the first graph if you want a permanent record, or rename it for temporary storage: not doing so will add the new graphs to the current plot window. To save, have the graph in focus, and click on the save button (fourth icon from the left on the 'Icon line') which should be highlighted, or click on `File` on the 'Command line' then on `Save as`. The default name offered will be 'Data Plot', and the default format will be .gwg (for GiveWin Graphics) a format from which *OxMetrics* can recreate the graph later. To rename the graph, again click on `File` on the 'Command line' then on `Rename` and enter the desired name in the box offered. The next task assumes you either saved or renamed the graph.

1.17 Task 2: Graphs with regressions

The second task will be to replicate Figure 1.2 and the following graphs of the logs of wages and prices, 'w' and 'p', through to Figure 1.4. The first steps are just like Section 1.16, but select w and p, then `Actual series`. Double click on the graph to bring up the `Edit Graph` dialog. Seven lines down is the entry `Regression, Scale` for w: click on that to reveal options for `Scaling` (which we will not need here) and `Regression`, where opposite `No of lines` change the 0 to 1, and just below that, `Type` will be highlighted. To see the deviations between the regression line and the data series, change `Sequential` to `Seq. with projections`. Click `Done` and a version of Figure 1.3 should appear. To more closely match, you will need to extend the range of the y-axis: to do so, bring up `Edit Graph`, and at the top, click on `Area 1 layout` to reveal three choices, of which you need to click on `World coordinates`. Select `Set Y`, and reset `Bottom` from -0.2 to -1.0, then `Done`.

Next bring up `Edit Graph` again, and reset 1 regression for w to 6; and the same for p (remember to set `Seq. with projections`). Now you are nearly at Figure 1.4: all you need to do is sign your picture. This time, **right** click with the mouse to bring up a small dialog, of which one choice is `Draw`: move the mouse over that entry to list all the drawing options, and left click on `Draw a Freehand line`. When the mouse is over the graph, you will see it has transformed into a pen. This pen can write whatever you wish when the **left** mouse button is held down, like a normal pen on paper, but with the *caveat* that you must not stop holding down the left mouse button (i.e., don't take the pen off the paper), but write continuously–which can take a little practice. Once you have written your signature, bring up `Edit Graph`, and a third line called x has appeared: this is your writing. Select 1 regression and `Seq. with projections` to fit a regression to your signature.[2]

1.18 Task 3: Scatter plots

The procedures for replicating Figure 1.5–Figure 1.7 are similar, so we move to replicating Figure 1.8. This is a new kind of graph called a scatter plot. Save, rename or delete your last plot to get a new graphics window, remove the current selection of 'w' and 'p', and select 'gl'= $g - l$ and 'kl'= $k - l$, but now click `All plot types` (this graph could be done by `Scatter plot (YX)` but it will be useful to learn the more general approach). The new dialog shows the wide range of types of graphs *OxMetrics* offers, and the many forms for each type. Here, click on `Scatter plots`, and in the right panel click on `Regression`, tick `Regression line` and also `With projections`, then `Plot` and `Cancel` to exit and see your 'production function'. Other choices of fitted lines (not necessarily linear) could be used by selecting `Smoothing` and some of the options that offers.

Return to the graphics icon and add *year* to the selection, `All plot types`, `3-D(XYZ) plots`, scroll down in the left-hand panel to find `3D points`, then `Plot` and `Cancel` to see a version of Figure 1.9 (which had the variables in a different order, and also rotated the graph for convenience, which can be done using the right mouse button, and selecting `Rotate`).[3]

The following few graphs should now be easy.

---

[2] This is less magical than it looks. The 'pen' available for writing on graphs in *OxMetrics* records the pixels where writing occurred, and the software knows the mapping between the units of the variables in a graph and their pixels, so can translate the writing into 'real world' magnitudes—and hence can fit a trend line to the writing, even including projections from that line. One can do lots of fun things with modern econometrics software...

[3] When a graph is moved, *OxMetrics* fixes its location. Such a move may be accidental, but then leads to added graphs partially overlapping. To correct the problem, double click on the apparently fixed graph, select `Area layout`, then `Pixel coordinates`, and untick the box opposite `Set` if it is ticked: if that box is blank, then try another area.

1.19 Task 4: Data density plots

The next new type of graph is to draw data densities, like those shown in Figure 1.14, although we will only explain the process for one of the sub periods. Click the graphics icon, clear the entry, and select 'w', 'p', 'g', and 'l', `All plot types` then `Distribution`. Tick `Normal reference` (options for `Histogram` and `Density` should be ticked already), and click on `Sample`, set `Sample start` to 1860 and `Sample end` to 1940, `OK`, then `Plot` and `Cancel`. Using `Edit Graph`, for each `Area 1-4` in turn, select `Histogram`, `Inside color`, and select grey (line 15 in the default settings).

Repeat for 1941 to 2011 (set the second date before the first to avoid zero observations). There should be 8 graphs, four shaded for the first sample and four unshaded for the second. Now, yet another special feature of *OxMetrics*: click on `Area 5` (the second sample density for w) so the jagged line appears round it, copy just that graph to the clipboard (or right click and select `copy`); click on `Area 1` (the first sample density for w, so now has the jagged line), and click `paste` to see the changes in the density between the periods. Repeat for each of the variables' matching subsamples, then delete the bottom four graphs (click to see the jagged line then right click selecting `delete`: hint–start with the last graph).

Hopefully graphing should now be straightforward, although there are a number of other types of graph used below, and often some additional 'tidying up' may be required to create presentational quality.

1.20 Chapter 1 exercises

1. Do a scatter plot of the data for the logarithm of wages, 'w', against the logarithm of prices, 'p'. Change the line `Type` from `symbols` to `line and symbols` and briefly discuss.
2. Fit a regression line with projections and briefly discuss.
3. Change to four regression lines with projections and discuss your set of findings.
4. Draw the data densities for 'wp', 'Ur', 'gl' and 'kl' in separate graph windows, like Figure 1.15, and discuss your findings in relation to that Figure.
5. Repeat Task 4 for 'wp', 'Ur', 'gl' and 'kl' by plotting their densities separately for 1860–1940 and then 1941–2011 and discuss the outcome in relation to Figure 1.15, explaining why some densities shifted considerably yet not all changed greatly (see Figure 1.17).

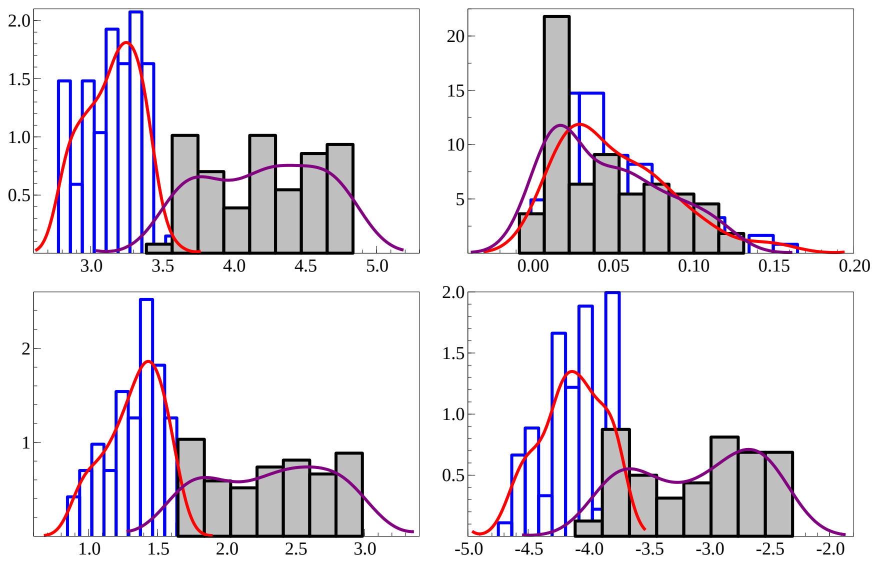


Figure 1.17
Histograms and densities of $(w-p)_t$, $U_{r,t}$, $(g-l)_t$ and $(k-l)_t$, shaded post 1940.

Chapter 2
Taming trends and breaks

Chapter 2 guide posts

1. While the levels of wages and prices trended greatly over the last 150 years, their rates of change did not trend: Section 2.1.
2. However, wage and price inflation experienced many sudden large changes in their average values at different point in time, called location shifts: Section 2.2.
3. Although the growth of real wages did not trend, it doubled after the Second World War, probably due to the growth of labour and capital: Section 2.3.
4. Most of the location shifts in wage and price inflation match, so real wages rarely jumped, called co-breaking, with the exception of a spike in 1940: Section 2.4.
5. Year-on-year changes in the log of wages divided by prices do not trend, so 'tame' the huge rises in the nominal levels of wages and prices.
6. Those huge rises are also tamed by scaling real wages by productivity per worker as that cancels their trends (called cointegration): Sections 2.5–2.7.
7. Indicator variables, which are zero everywhere except for unity over a short period to indicate its presence, are introduced to capture location shifts, and lead to an understanding of co-breaking: Sections 2.8–2.10.
8. Rapid shifts could also be due to non-linear reactions, so those are considered as well: Sections 2.11 and 2.12.
9. To summarize, as trends and breaks are ubiquitous in macroeconomics, we discuss how to handle both in two ways, where the first 'removes' the problem whereas the second 'models' it, namely differencing and cointegration for trends, and indicators and co-breaking for shifts.

2.1 Measuring wage and price inflation

Let us return to Victorian London in 1860, and again imagine you are living there on a worker's wage. At that time, the UK operated under the 'Gold Standard', which fixed the price of gold in terms of £. Gold had been discovered in California in the mid 1840s and in Australia in the late 1850s, then in the Klondike in the late

1890s, leading to general price rises for a time, but these were insufficient to sustain the rapid expansion of economies. When gold was scarce with no new discoveries to match the increases in output of goods and services, the price level had to fall, increasing the burden of debt on the poor: see Marshall (1926) in his evidence to the *Royal Commission on the Values of Gold and Silver*, 1888.

If we zoom in to the behaviour of wages and prices 1860–1900 as in Figure 2.1, this pattern shows up. Nominal wages and prices rose till the early 1870s, and then fell back till the late 1880s (panel a), a period when real wages lagged productivity growth (c). Wage and price inflation (b) fluctuated considerably, and both were often negative, so real wage changes (d) also varied between ±3%pa. Over your working life of 40 years, the purchasing power of your wage would have risen less than 40%–it started at an inadequate level, and ended at one. Many older individuals were forced into Workhouses, with much new construction in the 1860s.

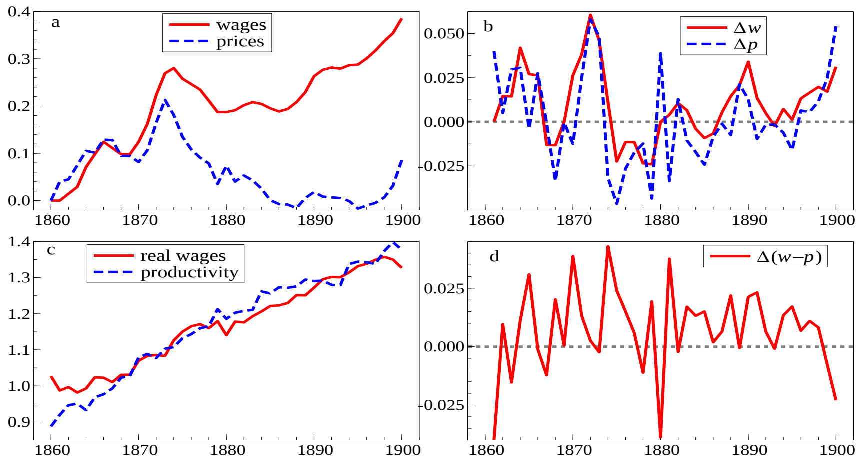


Figure 2.1
(a) Annual wages and prices; (b) wage and price changes; (c) real wages and productivity; (d) changes in real wages, all over 1860–1900.

Even by 1913, life was little better. In *Round about a Pound a Week* written by Maud Pember Pember Reeves (1913), she spells out the hardships of London life for many living well above the lowest poverty rung. The Great Depression did not help either, but during and after the Second World War, matters improved rapidly with the abandonment of the Gold Standard in favour of the Bretton Woods currency agreements. That collapsed in the 1970s, to be replaced by floating exchange rates, which necessitated some other mechanism to determine the price level.

Over the last decade or so, 'inflation targeting' has become the main instrument of economic policy in many countries, including the UK. The Bank of England adjusts interest rates to try and stabilize the rate of price inflation at around 2.0% p.a., measured by the consumer price index (CPI). We are analyzing the gross domestic product (GDP) deflator, P_t, an implicit price index that converts nominal GDP to real, but seeks to represent a general measure of prices entering GDP. Here, P_t is annual, whereas CPI is measured monthly. Thus, $\Delta p_t = p_t - p_{t-1}$ is annual inflation, and when $\mathrm{cpi}_t = \log(\mathrm{CPI}_t)$, so is $\mathrm{cpi}_t - \mathrm{cpi}_{t-12} = \Delta_{12}\mathrm{cpi}_t$. Price indices have arbitrary units determined by a 'base year' where they are set to unity, or sometimes 100. In logs, that affects the level, but not the changes.

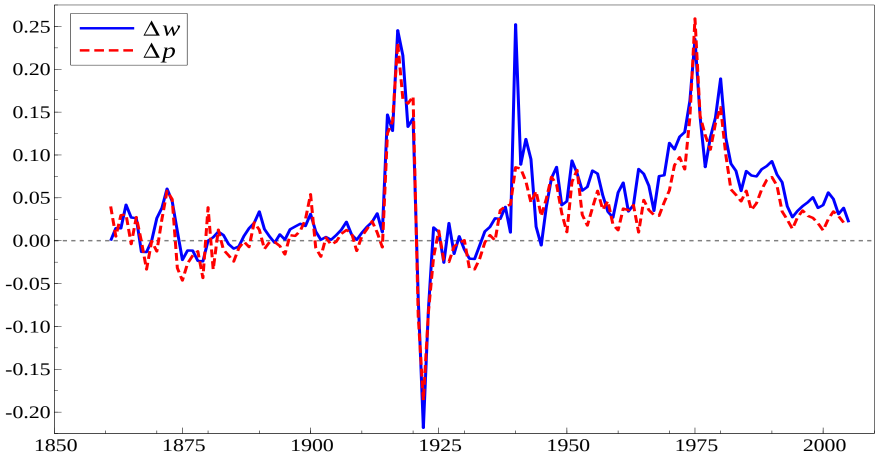


Figure 2.2
Annual wage and price inflation over 1860–2011.

Annual wage and price inflation, denoted Δw_t and Δp_t, are shown in Figure 2.2. The next section will analyze these transformations of the original time series.

2.2 Wage and price inflation outcomes

Figure 2.2 reveals many new features of these changes, which we list as follows:

1. neither Δw_t nor Δp_t trends overall, but both wander widely, over approximately ±25% p.a., consistent with one implication of a model like (1.14) when $\beta = 0$;
2. there are three episodes of very high inflation (defined as above 20% p.a.):
   a. for both Δw_t and Δp_t, during World War I, when wage indexation had been introduced, and lasting till about 1920;
   b. at the start of World War II for wages; and
   c. during the 1970s for both Δw_t and Δp_t, which was also a period with various incomes and prices policies to try and control inflation;
3. there are also periods of negative inflation, corresponding to sharp falls in wages and prices, especially over 1921–23, when both fell by up to 20%;
4. usually both variables move in similar ways, but;
5. generally, $\Delta w_t \geq \Delta p_t$, especially after WWII;
6. both series are relatively erratic with many 'jumps'.

Given this list, can you guess what the change in real wages, $\Delta(w_t - p_t)$, looks like? Figure 2.3, which also plots the sub-sample means pre- and post- 1945, shows there are some surprises.

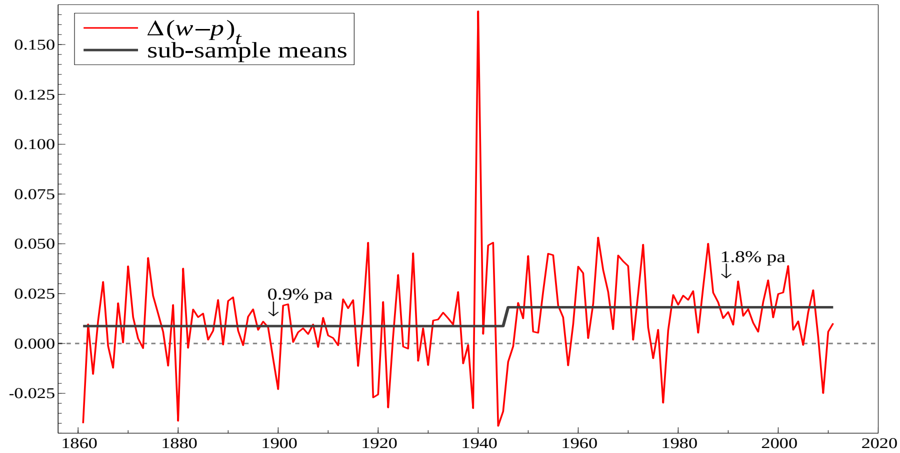


Figure 2.3
Real-wage growth over 1860–2011 with sub-sample means.

There is a huge spike in 1940, presumably to encourage a high participation in the labour force at the start of WWII: otherwise the range is from about −3% to about +5%, a much smaller range than the nominal variables. One can also see that the 19th Century had many movements, on a similar scale to the later data, despite the initial impression in Figure 1.1. The latter half of the 20th Century had a higher mean, with more persistent increases in the last few decades. However, it is hard to discern the small mean shift against the noisy fluctuations in $\Delta(w-p)$, even though the growth rate doubled from 0.9% p.a. pre-1945 to 1.8% p.a. after.

Such an apparently small shift can have a major effect on living standards. Real incomes double in 36 years at 2% p.a., so would grow 8-fold in a little over a century, and more than 16-fold over our period of 150 years. However, at 0.9% p.a. they only double every 80 years, so would be just 4-fold higher in 160 years (based on the so-called rule of 72). The growth rate of GDP per worker over 25-year intervals is shown in Figure 2.4 and highlights that UK prosperity is remarkably recent. Had the whole sample seen growth at 2% p.a. per person, real incomes would be more than twice as high as present levels.

Given the concomitant advances in hygiene, medicine and longevity, it is clear that we are very much better off now than 1860 on all such measures: but that does not explain why real wages (or output per person) increased so greatly. To unravel that, using the analysis in (1.6) that $(g-l)_t$ is mainly determined by $(k-l)_t$, we turn to examine the behaviour of the two key inputs to the production process, namely labour and capital.

2.3 Growth of labour and capital

The very different behaviour of Δl_t and Δk_t is shown in Figure 2.5, using the same units for the two graphs.

Employment has grown on average by 0.6% p.a., whereas capital has grown at 2% p.a., about 3.5 times as quickly. As a consequence, there is now vastly more

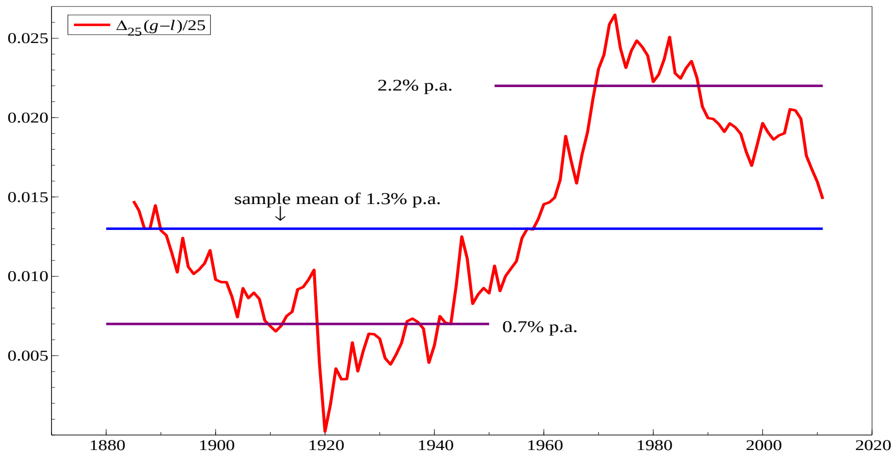


Figure 2.4
Real annual GDP growth per worker by 25-year intervals.

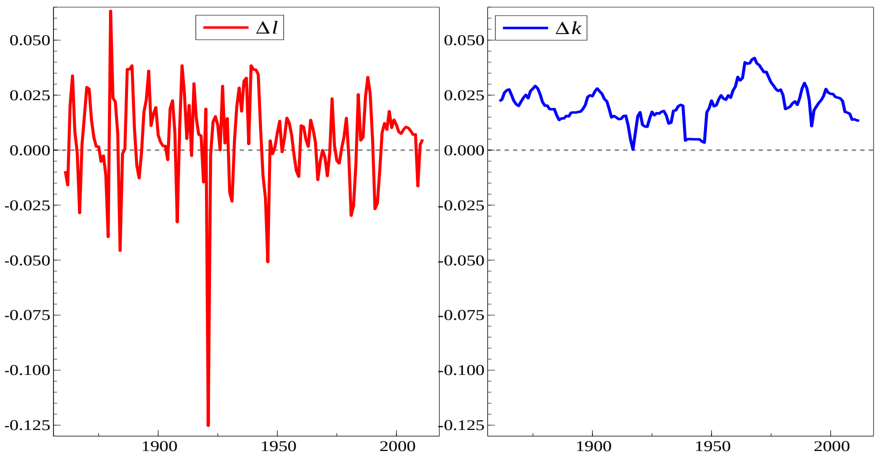


Figure 2.5
Growth rates of employment and capital stock over 1860–2011.

capital per worker, enabling workers to be much more productive. Moreover, each unit of capital is itself much more productive as it embodies more advanced technology, and the workers using that capital have much more education and greater skills, so higher human capital. Thus, although labour works much shorter hours per week, and many fewer hours per year with longer paid holidays, there has still been a huge rise in output per person employed as Figure 1.11 revealed. Despite growing more slowly than capital, employment has nevertheless been much more volatile: unfortunately it is easier to sack workers than equipment or buildings.

2.4 Taming time series

Do you remember Figure 1.1 in the left panel? J-shaped over [1, 700], with most early detail lost.

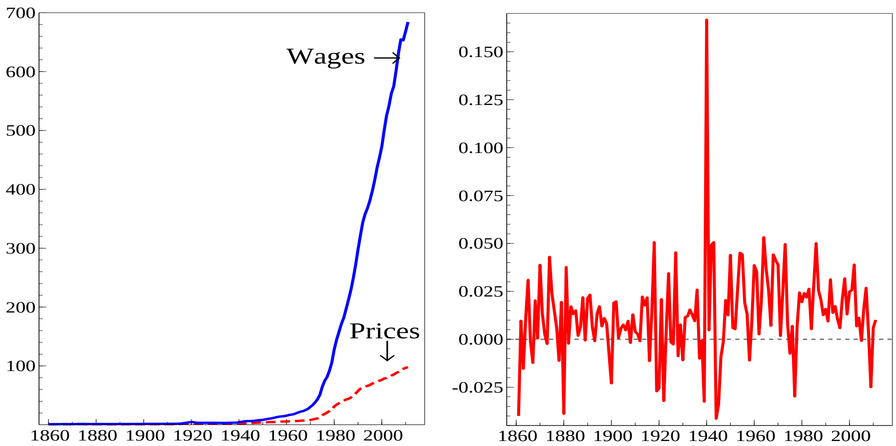


Figure 2.6
Taming trends in wages.

Figure 2.6 (right) shows considerable movement, but the behaviour is relatively similar over the 150 years, with $\Delta(w_t - p_t) \in [-0.03, 0.16]$ even when including the WWII spike. By taking logs of wages and prices, then the difference between those, and finally calculating the changes in that series, we have tamed the original dramatic trends. Nevertheless, the transformations allow us to recreate W_t given P_t from $\Delta(w_t - p_t)$, using:

$$W_t = \exp\left(p_t + \Delta(w_t - p_t) + w_{t-1} - p_{t-1}\right)$$

That is our first step in 'modelling' wages given prices.

2.5 Trend cancellation

There is another way to tame time-series trends, illustrated in Figure 2.7 which records the (log) 'wage share', $w_t - p_t - g_t + l_t$.

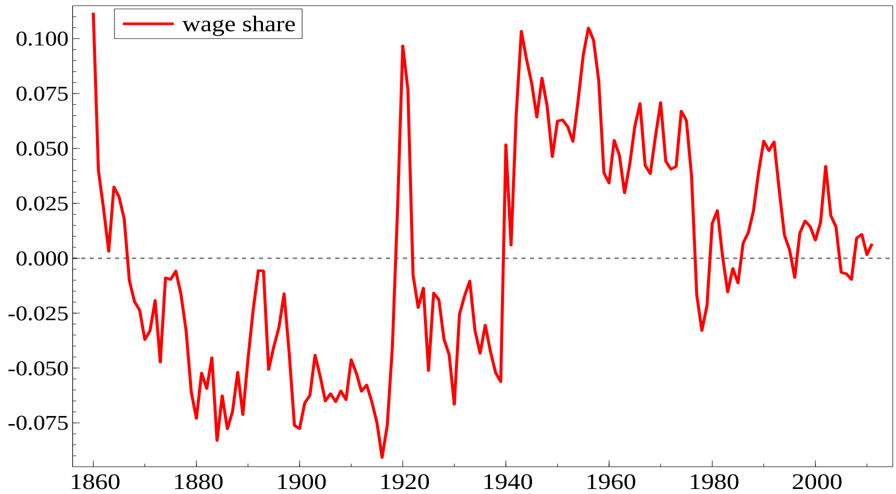


Figure 2.7
'Wage share' $(w - p - g + l)_t$ over 1860–2011.

Both $(w - p)_t$ and $(g - l)_t$ are cumulative processes with apparent stochastic trends, as Figure 1.11 showed. The 'wage share', $(w-p-g+l)_t$, is the gap between these, yet does not trend, demonstrating that a combination of variables can also cancel their trends.

As shown in section 1.11.1, time series with stochastic trends cumulate past shocks:

$$x_t = \alpha + \beta t + \sum_{s=1}^{t} \epsilon_s \tag{2.1}$$

An integral used to be written by an elongated S mimicking a sum, so time series like (2.1) are called integrated (here of first order), and denoted by I(1). Linear combinations of I(1) processes usually also exhibit stochastic trends: for example, g_t and l_t are I(1), and so is $g_t - l_t$. However, the cumulated shocks may cancel between time series, and cointegration is the name for such a property. Here, both $w_t - p_t$ and $g_t - l_t$ are I(1), but $w_t - p_t - g_t + l_t$ is not integrated, denoted by I(0).

2.6 Cointegration

Consider the two equations:

$$\begin{aligned} y_t &= \mu_0 + \gamma x_t + e_t \\ z_t &= \mu_1 + x_t + \nu_t \end{aligned} \tag{2.2}$$

where $\gamma \neq 0$, e_t and ν_t are random errors, and x_t is a stochastic trend:

$$x_t = x_{t-1} + \beta + \epsilon_t = x_0 + \beta t + \sum_{s=1}^{t} \epsilon_s \tag{2.3}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ as in (1.13). Then y_t and z_t 'inherit' a common stochastic trend from x_t so are I(1), but:

$$y_t - \gamma z_t = \mu_0 - \gamma\mu_1 + e_t - \gamma\nu_t \tag{2.4}$$

cancels the stochastic trend x_t by the (unique) linear combination $y_t - \gamma z_t$, which only depends on I(0) errors. While (2.3)–(2.4) are a simple example, they illustrate the general idea of cointegration.

2.6.1 'Natural' cointegration

Alternatively, two variables can be directly linked by:

$$y_t = \beta_0 + \beta_1 z_t + \epsilon_t \tag{2.5}$$

where ϵ_t is a random error, and z_t is I(1), generated by:

$$z_t = z_{t-1} + \delta + \nu_t \tag{2.6}$$

where ν_t is a random error. Then we can show that y_t is I(1) and cointegrated with z_t as follows.

Difference (2.5):

$$\Delta y_t = \beta_1 \Delta z_t + \Delta\epsilon_t \tag{2.7}$$

then using (2.6) implies:

$$\Delta y_t = \beta_1\delta + \beta_1\nu_t + \Delta\epsilon_t \tag{2.8}$$

Consequently:

$$y_t = y_{t-1} + \theta + u_t \tag{2.9}$$

where $\theta = \beta_1\delta$ and $u_t = \Delta\epsilon_t + \beta_1\nu_t$, so y_t is I(1). As $y_t - \beta_1 z_t = \beta_0 + \epsilon_t$ is I(0) from (2.5), and z_t is I(1), then y_t is cointegrated with z_t.

2.6.2 Multiple cointegration

Cointegration may require several variables, $y_t,\ z_t,\ w_t$, say, such that a linear combination $y_t + \lambda_1 z_t + \lambda_2 w_t$ cointegrates, where $\lambda_1 \neq 0,\ \lambda_2 \neq 0$, whereas no pair does. Then that triple is a minimal cointegrating set. The triple combination can be written as y_t cointegrating with $\lambda_1 z_t + \lambda_2 w_t$ or as z_t cointegrating with $y_t,\ w_t$ etc., and the normalization of which variable has a unit coefficient is often based on theoretical reasoning about the links.

Another useful property of cointegration is that when a group of variables cointegrates, they will still do so after adding a further variable, x_t. However, there may be another cointegrating combination involving some of the $y_t,\ z_t,\ w_t$ and x_t, say w_t and x_t. In turn, that must imply that $y_t,\ z_t,\ x_t$ also cointegrate, and again theory is needed to decide which set to choose.

Above, we saw that the wage share, $w_t - p_t - g_t + l_t$, was not integrated although the two pairs $w_t - p_t$ and $g_t - l_t$ were integrated. Consequently, it must be the case that any one of the four must cointegrate with a combination of the other three. The most economically interesting is between p_t and $c_t = w_t + l_t - g_t$, where c_t denotes unit labour costs, a combination of importance in Section 6.4.

2.7 Cointegrated time series

Many groups of economic time series seem to cointegrate. Figure 2.8 (left panel) shows the £ sterling exchange rate index, denoted e_t in logs, the logs of the UK price level, p_t, and world prices, $p_{w,t}$ (middle panel), and the resulting derived real exchange rate, or purchasing power parity, $ppp_t = e_t + p_t - p_{w,t}$ (right panel). Despite the opposite trends in the first two panels, ppp_t does not trend, and is little changed over 150 years, albeit having wandered widely.

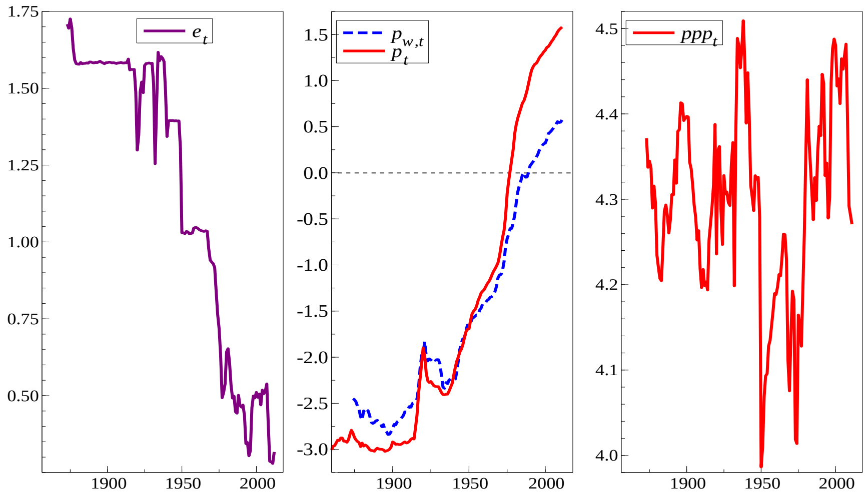


Figure 2.8
Exchange rate, prices and purchasing power parity, ppp_t.

A cointegrated relation defines an 'equilibrium trajectory', where departures induce an equilibrium correction that moves the economy back towards that path–otherwise the economy would drift–but convergence may be very slow. When ppp_t has been well above or below the average shown, it has tended to move back towards that average. 1965–1975 is an exception, where the £ fell sharply despite being well below parity.

As another example, Figure 2.9 shows the UK's long-term and short-term interest rates, R_L and R_S, and their spread, $R_L - R_S$. The long rate acts like a long moving average of past short rates, although marked departures can occur, as from 1930–1955, and again after 2008. Overall, the spread is centered near zero, with a range of about ±3%. We will return to cointegration analysis in Task 8 Section 2.17, and the topic will also recur later.

2.8 Evidence about breaks in nominal and real wage growth

The next step is to investigate breaks in nominal and real wage inflation over the last 150 years. Table 2.1 records the values of their means and standard deviations (SDs) over the whole period, and sub-samples pre- and post-WWII, all expressed as percentages (SD is in the same units as the variables when logs are used).

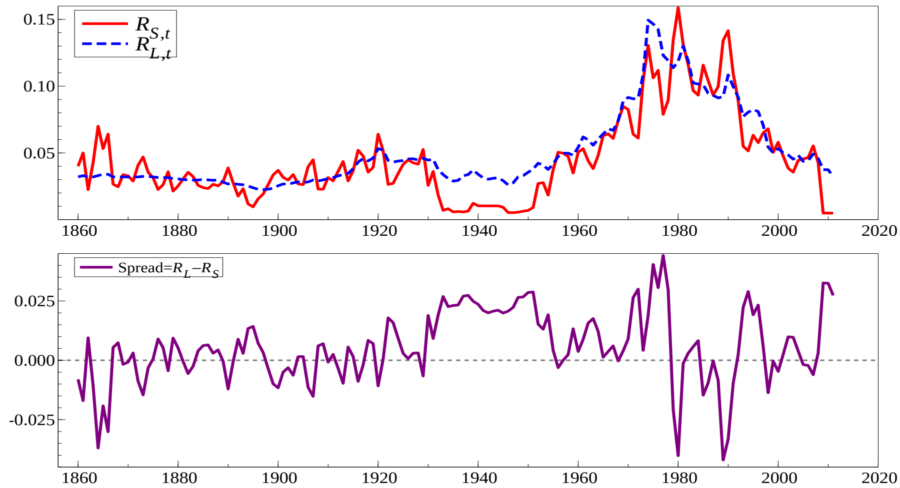


Figure 2.9
Interest rates and spread over 1860–2011.

The means of Δw_t and $\Delta(w-p)_t$ are much higher post-1945, yet both those variables must have been more stable as they have smaller SDs post WWII. Also, SD$[\Delta(w-p)]$ is always much smaller than SD$[\Delta w]$. Consequently, the whole-period mean and SD are not a good representation of what happened, matching the graphs of densities in Figure 1.15.

1945 was chosen as the split year because a number of major social structural changes occurred during 1945–46 or shortly thereafter, including:

- nationalization of major basic industries like coal and steel;
- the introduction of the National Health Service (NHS);
- large changes to unemployment insurance and pensions;
- the Beveridge (1942), Beveridge (1945) reports gave a mandate for seeking low national unemployment using Keynesian policies.

mean (SD)	1861–2005	pre-1945	post-1945
Δw_t	4.3% (5.9%)	2.2% (6.2%)	7.1% (4.2%)
$\Delta(w-p)_t$	1.3% (2.3%)	0.9% (2.6%)	1.8% (1.7%)

Table 2.1
Means and standard deviations of Δw_t and $\Delta(w-p)_t$ overall and two sub-periods in % p.a.

However, there were a number of other potential split dates over the period, corresponding to major institutional, political, or military changes, shown in Figure 2.10.

Matching these regime changes, the sub-sample mean values shown in Figure 2.11 suggests there were nine mean shifts in the level of nominal wage inflation. These nine distinct 'epochs' are marked as follows:

(1) a business-cycle era over 1865–1914, with 6 cycles of about 8 years each;

Figure 2.10
Wage inflation and institutional regimes.

(2) a huge rise in wage inflation during WWI, then;
(3) a very large crash over 1921–1923;
(4) wage inflation then remains near zero till the start of WWII;
(5) with a huge 'spike' in 1940 and a higher level during the war;
(6) then falls again at the end of the war, remaining low and steady till the mid-1970s;
(7) followed by a sharp rise during the oil crises;
(8) falling back after the UK leaves the European Exchange Rate Mechanism (ERM) in late 1992;
(9) finally low levels during the 'Great Recession' and quantitative easing.

The shifts in the means shown in Figure 2.11 are called location shifts. Table 2.2 records the numerical values of the sub-sample means. There are very large changes in both Δw_t and Δp_t over time, and relative to the present Bank of England price inflation target of 2.0%; the shift in the mean growth of $\Delta(w-p)_t$ pre- and post-WWII is also clear; as is the 'cancellation' of most of the large nominal changes in the real wage measure.

Large shifts in means (and variances) over time pose very different problems from deterministic and stochastic trends: such shifts are usually unanticipated, so can lead to large forecast or expectations errors, well illustrated by the recent financial crisis.

These sub-sample divisions were chosen after the fact based on the institutional regimes known historically, and from ocular inspection of the graphs of the time series. More formally, we need general procedures for detecting and dealing with shifts.

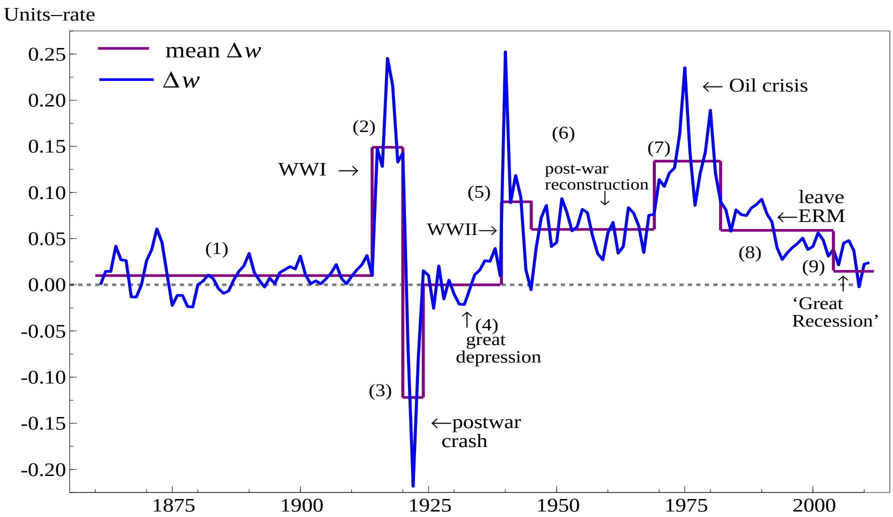


Figure 2.11
UK wage inflation location shifts over 1860–2011.

Epoch	Sub-sample	Δw_t	Δp_t	$\Delta(w-p)_t$
(1)	1861–1913	1.00	0.20	0.80
(2)	1914–1920	14.6	14.0	0.60
(3)	1921–1923	−12.2	−11.9	−0.30
(4)	1924–1938	0.50	−0.50	0.90
(5)	1939–1945	8.20	5.90	2.30
(6)	1946–1968	6.00	3.90	2.10
(7)	1969–1981	13.4	11.9	1.60
(8)	1982–2011	5.20	3.50	1.70
2005 – 2011	2.80	2.28	0.52	

Table 2.2
Mean shifts over nine sub-periods in % p.a.

2.9 Representing a single location shift

Consider a single mean shift from μ_0 to μ_1 at time $T_1 < T$ represented mathematically by:

$$y_t = \begin{cases} \mu_0 + \epsilon_t & t \leq T_1 \\ \mu_1 + \epsilon_t & t > T_1. \end{cases} \tag{2.10}$$

where ϵ_t is a random error with mean zero. The overall mean of y_t is a weighted average across the time spent in the two regimes:

$$\mathsf{E}\left[y_t\right] = \frac{1}{T}\left(\mu_0 \times T_1 + \mu_1 \times (T - T_1)\right) = \mu_0 + \frac{(T - T_1)}{T}\left(\mu_1 - \mu_0\right). \tag{2.11}$$

To model such a shift, we use an indicator variable, denoted by $1_{\{t>T_1\}}$, which is zero till T_1 then unity after:

$$\begin{aligned} 1_{\{t>T_1\}} &= 0 \quad t \leq T_1 \\ 1_{\{t>T_1\}} &= 1 \quad t > T_1 \end{aligned}$$

so we can write:

$$\begin{aligned} y_t &= \mu_0(1 - 1_{\{t>T_1\}}) + \mu_1 1_{\{t>T_1\}} + \epsilon_t \\ &= \mu_0 + (\mu_1 - \mu_0)\, 1_{\{t>T_1\}} + \epsilon_t \end{aligned} \tag{2.12}$$

Then (2.12) has an intercept of μ_0, with a location shift of $\mu_1 - \mu_0$ affecting data after time T_1. Figure 2.12 illustrates.

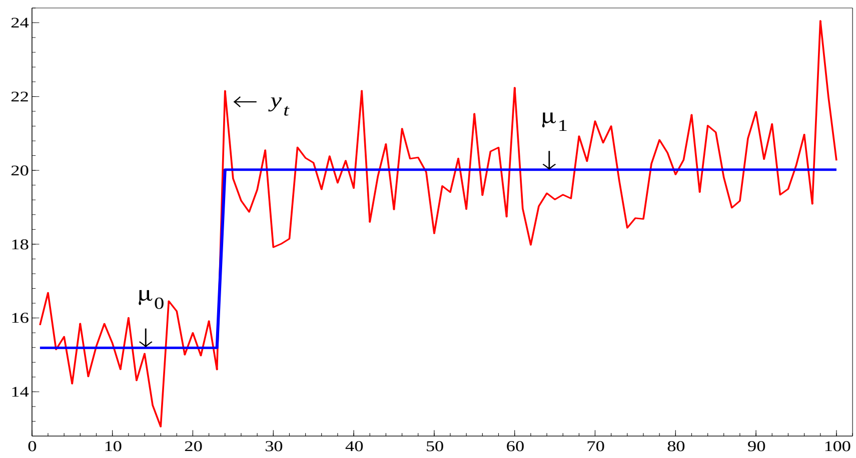


Figure 2.12
Location shift at $T_1 = 23$ from $\mu_0 = 15$ to $\mu_1 = 20$.

2.9.1 Location shift example

Figure 2.12 represents an artificial data series generated by (2.12) with an error $\epsilon_t \sim \mathsf{IN}[0, 1]$ where the shift is $\mu_1 - \mu_0 = 5$ standard errors occurring at time $T_1 = 0.23T = 23$ from $\mu_0 = 15$.

To represent such a shift here would need an indicator $1_{\{t\geq 23\}}$ for the second mean, $\mu_1 = 20$, as only later observations are affected. However, Figure 2.11 had eight mean shifts, so we must be able to model multiple location shifts.

2.10 Multiple shifts

To model many shifts, we use multiple indicator variables, $1_{\{1\leq t<T_0\}}$, $1_{\{T_1\leq t<T_2\}}$, $1_{\{T_3\leq t<T_4\}}$ each of which is unity over the period shown, so:

$$y_t = \mu_0 + (\mu_1 - \mu_0)\, 1_{\{T_1\leq t<T_2\}} + (\mu_2 - \mu_0)\, 1_{\{T_3\leq t<T_4\}} + \cdots + \epsilon_t \tag{2.13}$$

Now (2.13) has an intercept of μ_0, with a mean shifts of $\mu_1 - \mu_0$ at time T_1, $\mu_2 - \mu_0$ at time T_3, etc.

The eight mean shifts in Figure 2.11 are the differences between successive mean values in Table 2.2, namely 13.6%, −13.2%, −0.5%, 7.2%, 5.0%, 12.4%, 4.9% and 1.8% from an initial-sample mean of $\mu_0 = 1.0\%$. These could be represented by 9 indicators and no overall mean; or eight indicators and an overall mean; but not 9 indicators and an overall mean (check you understand why). To 'explain' such shifts would require some other variable that rose and fell in precisely those patterns and magnitudes at the same times. When breaks occur in several series, but a combination of them has fewer breaks, we call that co-breaking, as illustrated in the next section.

2.10.1 Graphical evidence of co-breaking

The graphs of wage and price inflation rates (Figure 2.13, left panel) show many more shifts than real-wage changes (Figure 2.13, right panel). This matches the phenomenon already noted in Table 2.2. We will now analyze co-breaking between Δw and Δp with many shifts, to produce $\Delta(w-p)$ with few.

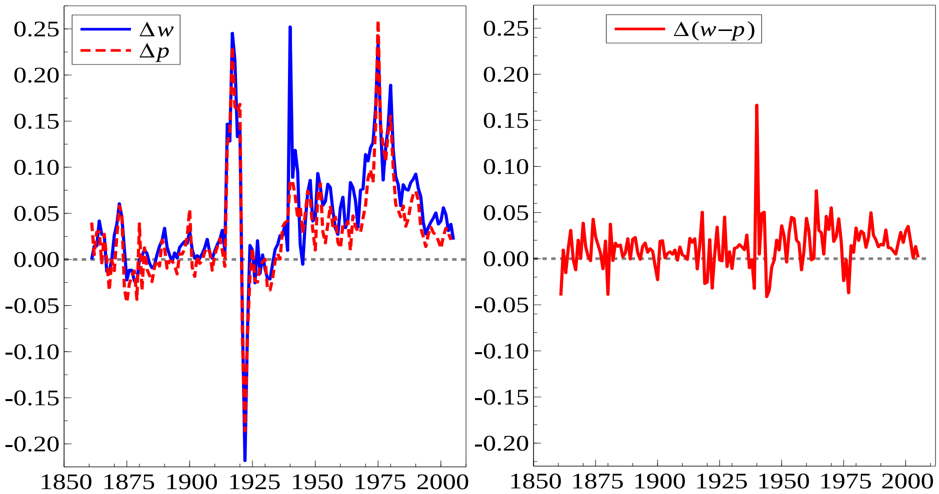


Figure 2.13
Evidence of real-wage co-breaking.

2.10.2 Location shifts and co-breaking

First, for a single break, consider the two equations:

$$\begin{aligned} y_t &= \mu_0 + \left(\mu_1 - \mu_0\right) 1_{\{t \geq T_1\}} + e_t \\ z_t &= \mu_0 + \left(\mu_1 - \mu_0\right) 1_{\{t \geq T_1\}} + \nu_t \end{aligned} \tag{2.14}$$

Then subtracting z_t from y_t:

$$y_t - z_t = e_t - \nu_t \tag{2.15}$$

so the shift has cancelled for that unique linear combination. While (2.14) is again a simple example, it illustrates co-breaking. Extending that theoretical analysis to multiple shifts as in (2.13) is straightforward, though having all shifts cancel is demanding empirically.

2.11 Non-linear reactions

Instead of the linear relations of the form $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$ so far assumed, y_t may be a non-linear function of x_t denoted $y_t = f(x_t) + \epsilon_t$. There is a vast number of possible non-linear functions, the simplest of which is a polynomial of the form:

$$y_t = \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \theta_3 x_t^3 + v_t \tag{2.16}$$

where further powers of x_t could be included. One advantage of the formulation in (2.16) is that despite being non-linear in the variable x_t, it remains linear in the parameters θ_i so can be estimated by regression methods: to see that it is really a linear model, redefine $x_t^i = z_{i,t}$ for $i = 1, 2, 3$. A second advantage is that (2.16) can be used to test whether $f(x_t)$ is non-linear by testing the significance of x_t^2 and x_t^3, an example of the test for non-linearity in Castle and Hendry (2010). Their approach was to compute orthogonal linear combinations, $u_{i,t}$, of the original variables (called principal components), and test the significance of adding squares, cubes and exponential functions of each individual component to a linear model.

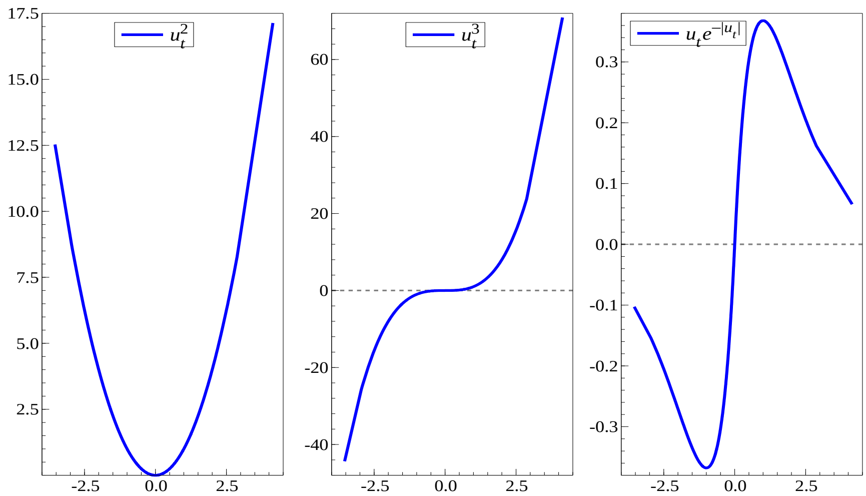


Figure 2.14
Non-linear transformations

For a single regressor, that idea simplifies to using the powers of x_t. However, a disadvantage of (2.16) is that the form of $f(x_t)$ in the data generation process (DGP, which is the 'true' process) is rarely known, so a polynomial is likely to be just an approximation. A second disadvantage is that polynomial terms are potentially unbounded, and cubics can become large. Using deviations from means (denoted $\bar{x}$) can help standardize values, as in (2.17), but doing so changes the interpretation of other coefficients as (e.g.) $(x_t - \bar{x})^2 = \bar{x}^2 - 2\bar{x}x_t + x_t^2$. To capture bounded behaviour, we can also add the exponential term $x_t \exp(-|x_t|)$ which converges to zero for both very large and very small values of x_t, leading to:.

$$y_t = \phi_0 + \phi_1 x_t + \phi_2 (x_t - \bar{x})^2 + \phi_3 (x_t - \bar{x})^3 + \phi_4 (x_t - \bar{x}) \exp(-|x_t|) + v_t, \tag{2.17}$$

Figure 2.14 shows graphs of a quadratic (first column), cubic (second column) and the exponential term (third column), so a formulation like (2.17) can represent a variety of non-linear reactions.

A widely used non-linear model in macroeconomics is called a logistic smooth transition (see Maddala, 1977, Granger and Teräsvirta, 1993, and Teräsvirta, 1994).

The idea is to allow for more than one 'regime' between which the economy moves more or less rapidly, where both the number of regimes and the speeds of adjustment are determined from the data evidence. First, let $s_t = (x_t - c)/\sigma_x$, which is a scaled and recentered version of x_t, so s_t is measured in units of the standard deviation, σ_x, of x_t, where c is a threshold. Next, consider the function $\exp(-\gamma s_t)$ where $\gamma > 0$: as s_t becomes large and positive, the function converges to zero, whereas when $s_t < 0$, bigger negative values will lead the function to become unbounded. To ensure a bounded outcome, the transform $1/(1+\exp(-\gamma s_t))$ is used, which becomes unity for large positive s_t and zero for large negative values of s_t leading to the following formulation for a variable y_t as a function of x_t:

$$y_t = \mu + \beta x_t + \mu^* \left(1 + \exp\left(-\gamma s_t\right)\right)^{-1} + \epsilon_t \tag{2.18}$$

In (2.18), μ and β are the original mean and regression parameter, and $\mu + \mu^*$ is the mean after the shift to the second 'regime' is completed. That shift depends on how far x_t is from the threshold c, in units of σ_x, and $\gamma > 0$ determines the rapidity of the adjustment. Such a model is non-linear both in the variables (through the exponential function) and in the parameters. The latter is most easily seen by setting $\mu^* = 0$ so that γ disappears from the model, which raises what is called an identification problem as the value of γ is then indeterminate. That is in contrast to say (2.14), where even if the two means are equal so μ_1 disappears, that can be ascertained from the zero coefficient on the indicator $1_{\{t \geq T_1\}}$. Although there are solutions to the identification difficulty, it is very hard to select a model from data when any parameters enter non-linearly.

The formulation in (2.16) can be seen as an approximation to (2.18) based on approximating $1/(1+\exp\left(-\gamma s_t\right))$ by the first term of its power series:

$$\left(1 + \exp\left(-\gamma s_t\right)\right)^{-1} \approx 1 - \exp\left(-\gamma s_t\right) \tag{2.19}$$

then expanding the right-hand side exponential function as a polynomial:

$$\exp\left(z\right) \approx 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 \tag{2.20}$$

Since $-\gamma s_t$ is just a scaled version of x_t, (2.18) can be re-written as:

$$y_t \approx \theta_0 + \theta_1 x_t + \theta_2 x_t^2 + \theta_3 x_t^3 + v_t, \tag{2.21}$$

which becomes (2.17) on adding $\theta_4 x_t \exp(-|x_t|)$.

2.12 Location shifts and non-linear reactions

Although many aggregate economic time series appear to exhibit location shifts, sudden movements in levels could also be the outcome of a non-linear reaction. For example, Figure 2.15 shows the change in the own-interest rate on chequing accounts at UK Commercial Banks (R_S, scaled by 10 for readability) following the Finance Act of 1984, with both location shift and non-linear approximations. There was a sharp rise in R_S from zero to about 8.5%p.a., after which the rate levelled off. The location shift approximation captures the broad outlines of the movements, and would certainly be an improvement over not allowing for the change in R_S. The non-linear approximation uses an exponential function which is zero till

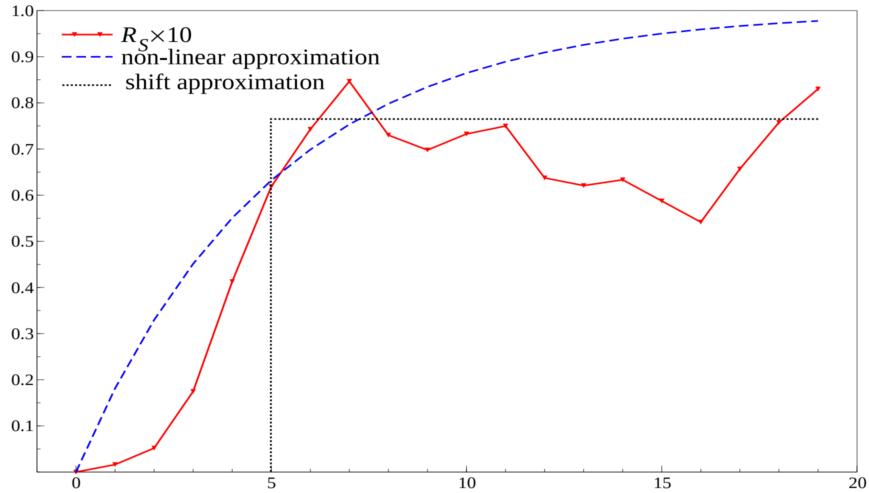


Figure 2.15
Own-interest rate on chequing accounts.

1984(2) then $(1 - \exp\{-.18(t - 1984(2))\})$, which fits better initially, but overshoots later.

Unfortunately, the wrong choice between a location shift and a non-linear reaction can have an adverse impact on forecasts by incorrectly extrapolating into a future period a non-existent shift, or a spurious non-linearity. Fortunately, when the DGP is in either the class of shifts or non-linearities, it seems that the model selection methods described in Chapter 4.9 will usually choose a useful approximation. When the jumps are due to the same non-linear reaction at different times, that offers a more parsimonious representation than many separate indicators, so is usually chosen; and when the shifts are not connected, indicators will usually replace an incorrect non-linear function. We will return to non-linear models in Chapter 6.3.1.

2.13 Chapter 2 key points

(A) Wage and price inflation did not trend over the last 150 years.
(B) Both show many sudden large location shifts.
(C) Many of those location shifts match, so real wages rarely jump, with the exception of the huge spike in 1940.
(D) Real-wage growth is trend free, but doubled after WWII.
(E) The rate of productivity growth also changed: capital became more productive, capital per worker grew, and workers were more educated and skilled.
(F) Real-wage growth (i.e., differencing the level of the log differential) 'tames' the huge rises in nominal levels of wages and prices.
(G) We can also 'tame' those huge rises in nominal levels using cointegration to cancel trends.
(H) Next, indicator variables were introduced to 'remove' breaks.
(I) That allowed the formulation of a model of co-breaking, cancelling breaks.

(J) However, rapid shifts may be due to non-linear reactions, so we also considered some non-linear representations.
Since trends and breaks are ubiquitous in macroeconomics, we have discussed how to handle both in two ways:
a] trends: by (i) differencing and (ii) cointegration;
b] breaks: by (i) indicators for shifts and (ii) co-breaking.
The first 'removes' the problem whereas the second 'models' it.

Before implementing such ideas, we will consider dependence between variables, and over time.

2.14 Task 5: Calculator operations

This task introduces the `Calculator` and its operations in *OxMetrics* to replicate Figure 2.4 reporting real annual GDP growth per worker by 25-year intervals. Click on the calculator icon (8th from the right on the Icon line) or type Alt+**C**, to bring up the `Calculator`. There is a window on the second row where operations will appear, the left-hand side shows typical calculator buttons, the centre offers some frequently used operators, and the right lists the `Database` variables. You will need to implement the formula $\left((g-l)_t-(g-l)_{t-25}\right)/25=\Delta_{25}(g-l)/25$. To do so, highlight gl, then click `diff...` to bring up the dialog for `Lag length` which shows 1, but you want to enter 25, then click `OK`. The calculator window should now show 'diff(gl, 25)', so change that to 'diff(gl, 25)/25' and click the large '=' button to implement. A suggested `Destination name` of 'D25gl' will appear, but change that to 'D25gl25' and click `OK` to accept and create. Exit the calculator.

OxMetrics uses a number of default naming conventions as follows. A leading **L**, as in **L***EmpUK*, denotes a log, although the database has renamed many variables to just the corresponding lowercase letter, so p rather than LP denotes $\log P$. A leading **D**, as in **D***p*, denotes a first difference, whereas $D2p$ would be a two-period difference ($p_t - p_{t-2}$) and DDp would be a second difference ($\Delta p_t - \Delta p_{t-1} = \Delta^2 p_t$. A lag in models is denoted G_1 for G lagged one period, and that is also the `Destination name` offered if the calculator creates a lag. Names can include such symbols or operators, so $w-p$ is possible, but to distinguish such names from the operation $w-p$, the former need to appear in double quote marks as "$w-p$": similarly for G_1, so it may be easier to use $G1$ where the last number denotes the lag length.

2.15 Task 6: Algebra operations

This task extends Task 5 Section 2.14 to `Algebra` operations. Click on the `Results` name under `Text` in the far left column of the *OxMetrics* window to see:
Algebra code for UKHist2013.xls:
D25gl25 = diff(gl, 25)/25;
`Calculator` operations are written to the `Results` file, so can be collected and saved in an `Algebra` file (with extension .alg) as follows. Highlight the operation 'D25gl25 = diff(gl, 25)/25;' by dragging the mouse over it, click the copy button and then click on the algebra icon `Alg` (also Alt+**A**) to bring up the `Algebra` dialog. Right click with the mouse to enter the operation on line 1, then `Save as` (with

the desired name). Note that algebra commands are always terminated by a ';' to denote where one command ends and the next starts.

The results file is 'live' in the sense that highlighted operations can be executed. If 'D25gl25 = diff(gl, 25)/25;' is highlighted, Ctrl+**A** will implement that operation. An important difference between the `Calculator` and `Algebra` is that the former checks if you want to overwrite a variable with the same name as one already in the database, but `Algebra` does not. Thus, good practice is never to save an altered 'core' database, but save all the required transformations in an `Algebra` file and run it at the start of a new session.

2.16 Task 7: Documenting and modifying graphs

Next, graph the variable 'D25gl25' (newly created variables are added at the end of the database). To rename the legend, select `Y label` and type in:
$\Delta _{25}(g-l)/25$.
To draw the lines, right click the mouse and select `Draw`, `Draw a straight line`, and use the pen to draw in the region of 0.007 from about 1880 to 1950, finishing by pressing the `Ctrl` key on the keyboard, which makes the line straight. Double click on that line (under `Lines and symbols`), select `Set world coordinates`, and set both `Bottom` and `Top` to 0.007, and if needed, `Left` to 1880 and `Right` to 1950. Similarly for the other lines and also those in Figure 2.10.

A useful variation allows us to handle figures with multiple graphs, like Figure 2.5. In the `Graphics` dialog, select 'Dl' and 'Dk', then on `Actual series (separately)`, which will produce two graphs stacked vertically. First, in `Edit Graph`, select `Graph layout`, then `Areas`, changing `2x1` to `1x2`, so the two graphs will appear beside each other (click `Apply` to check). Now click on `Area 1` in the left column, and at the foot of the right column, click `Copy properties to other areas`. To align the scales of the two growth rates, select `World coordinates` and tick the '`Y`' box, then `Done`. A similar procedure applies for (say) Figure 2.8.

2.17 Task 8: Deterministic versus stochastic trends

The final Task for this chapter is to use graphs to investigate deterministic as against stochastic trends. First graph g (the log of GDP), and select one regression line with projections to see panel a in Figure 2.16.

A single deterministic trend creates very substantial and systematic deviations from the actual data. In comparison, fitting a single regression line with projections to Δg as shown in panel b still creates substantial deviations, but they are much less systematic. One way of emphasizing the difference between deterministic and stochastic trends is to fit 6 regression lines to each of g and Δg, as shown in panel c and d. There is a marked reduction in the magnitude and systematic nature of the deviations for g, but almost no change for Δg.

To interpret these findings, reconsider the model in (1.12), expressed as in (1.14):

$$x_t = x_0 + \beta t + \sum_{r=1}^{t} \epsilon_r \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{2.22}$$

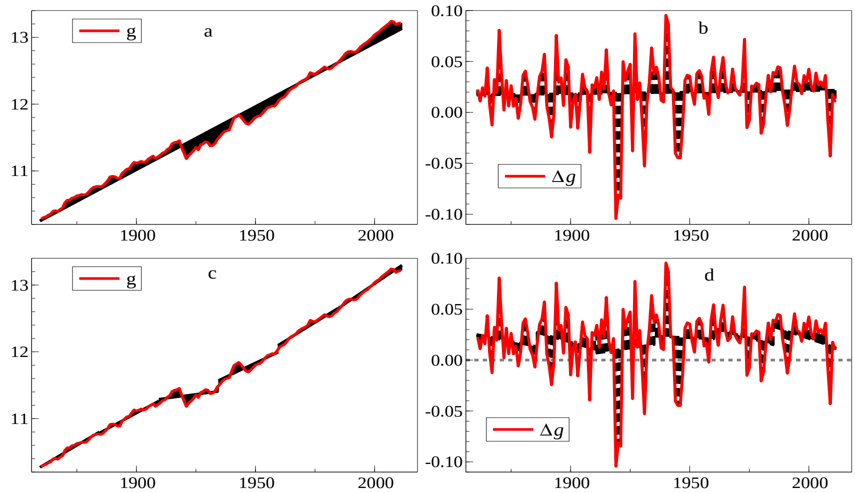


Figure 2.16
Deterministic versus stochastic trends.

Even if the trend was at a constant rate, as (2.22) assumes, the deviation from a linear trend would evolve slowly, from $\sum_{r=1}^{t-1} \epsilon_r$ to $\sum_{r=1}^{t} \epsilon_r$, so would look systematic. Conversely, Δx_t would only depend on ϵ_t, so the deviations would seem unsystematic. Neither model provides the complete story, as there appear to be changes in the rate of growth, seen most clearly in panel *c*; and the deviations seen in both *b* and *d* are not random.

2.18 Chapter 2 exercises

1. Do a scatter plot of nominal wage growth, 'Dw', against price inflation 'Dp'. Change the line `Type` from `symbols` to `line and symbols` and discuss what the time-linked lines show.
2. Fit a regression line with projections: does it change your interpretation in 1.?
3. Copy the graph in 1. and change to four regression lines with projections in the copy: can you explain why there is so little difference from 2.? Are six regression lines different?
4. Calculate the mean values of the growth of real wages $\Delta(w - p)$ before 1945 and again from 1946–2011 using `Descriptive Statistics using PcGive` in `Model class` (select 'Dwp', click `OK`, then choose `Means, standard deviations and correlations` which is probably already marked, `OK`, and set `Estimation ends` at 1945, then `OK`, repeating for the second sample).
5. Plot the time series for 'Dwp', as in Figure 2.13, right panel, and use the pen to draw straight lines at the two sub-sample means on the graph (see Task 7, Section 2.16). Had you noticed how different these means were?

Chapter 3
Dependence between variables and over time

Chapter 3 guide posts

1. Most macroeconomic variables are highly correlated, but their correlations are not constant over time: Section 3.1.
2. The correlations between the current levels of many economic variables and their previous values (called autocorrelations) are even higher, often being close to unity 20 years apart: Sections 3.2 and 3.3.
3. Such findings are consistent with macroeconomic variables having stochastic trends, but being perturbed by shifts: Section 3.4.
4. Matching that, autocorrelations between the differences of variables usually 'die out' rapidly as their distance apart increases: Section 3.5.
5. Correlations between differences of economic variables are often low, and indeed real-wage growth not only removes the stochastic trends of wages and prices, and most of their breaks, but also the autocorrelations of wage and price inflation: Section 3.6.
6. The relationship between unemployment and wage inflation known as the Phillips curve is shown to shift over time: Sections 3.7–3.9.
7. Non-constant relationships are common in empirical macroeconomics, suggesting important determinants have been omitted: Section 3.10.
8. To understand the behaviour of macroeconomic variables requires models that include all substantively relevant variables, their dynamic reactions, any non-linearities, all breaks, and any trends.

3.1 Interdependence between variables

Aggregate economic variables are highly intercorrelated, as Table 3.1 illustrates for the sample correlations between the five main time series graphed above, denoted

$\mathsf{corr}(x_t, y_t)$ for any pair x_t, y_t and defined by:

$$\mathsf{corr}\left(x_t, y_t\right) = \frac{\frac{1}{T-1}\sum_{t=1}^{T}\left(x_t - \bar{x}\right)\left(y_t - \bar{y}\right)}{\sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2 \frac{1}{T-1}\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2}} \tag{3.1}$$

where:

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t \text{ and } \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

1860–2011	p	w	g	l	k
p	1.000	**0.998**	**0.946**	0.819	**0.952**
w	–	1.000	**0.964**	0.848	**0.967**
g	–	–	1.000	**0.950**	**0.997**
l	–	–	–	1.000	**0.939**
k	–	–	–	–	1.000

Table 3.1
Sample correlations between macroeconomic time series.

All ten sample correlations are large and positive, and eight exceed 0.9, as shown in bold. Figure 3.1 records their time series as a group (together with the UK population, denoted *pop*) matched by means and ranges to visually illustrate their high correlations. Figure 3.2 shows a selection of scatter plots between these series.

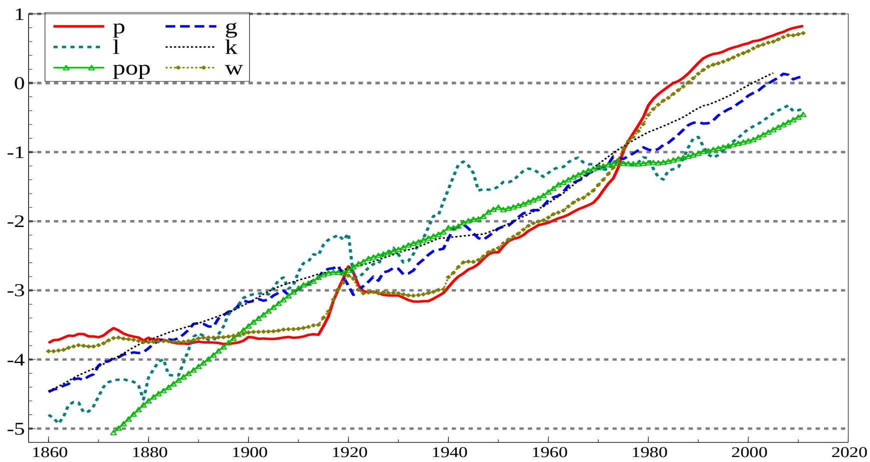


Figure 3.1
Many aggregate economic time series.

Six features of the interdependence shown by the scatter plots in Figure 3.2 are common in macroeconomic time series.

(1) such high correlations are quite unlike most cross-section scatter plots;

(2) points are plotted—but the outcomes look like lines;

(3) all the plots share the feature of going from lower left to upper right;

(4) however, most are far from straight lines, even in rows 1 and 2, where they are organized to be closest to linear;
(5) some rise sharply early, then ‘flatten off’ (row 3);
(6) others start ‘flat’, then suddenly rise (row 4).

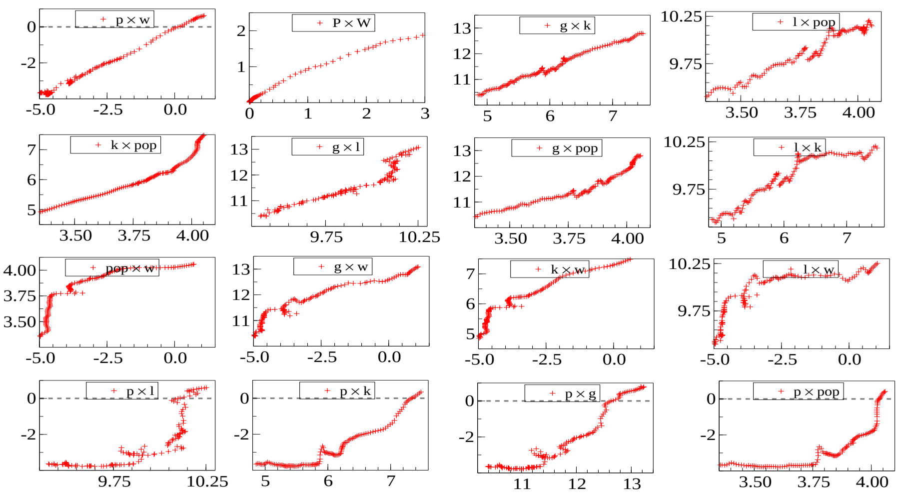


Figure 3.2
Scatter plots of economic variables.

Consequently, although the correlations are generally very high, they vary over time, so linear relationships are not constant. Moreover, many variables ‘proxy’ each other, potentially inducing ‘spurious relations’. However, a third difficulty lurks hidden behind these correlations between variables, namely there are high correlations between successive values of each variable, violating independent sampling within each series.

3.2 Time correlations over 1870–2011

We are now moving from investigating $\mathsf{corr}(x_t, y_t)$ to calculating correlations of the form $\mathsf{corr}(x_t, x_{t-1})$, called (sample) ‘autocorrelations’.

$\mathsf{corr}(x_t, x_{t-s})$	x_{t-1}	x_{t-2}	x_{t-3}	x_{t-4}	x_{t-5}
p_t	0.999	0.998	0.996	0.993	0.991
w_t	0.999	0.998	0.997	0.995	0.993
g_t	0.999	0.996	0.996	0.994	0.993
l_t	0.996	0.990	0.983	0.977	0.974
k_t	0.9999	0.9997	0.9994	0.9989	0.9983

Table 3.2
Sample autocorrelations for macroeconomic time series.

Table 3.2 calculates such autocorrelations for $s = 5$ lags. All are large and positive, and even the smallest is as large as 0.974, consistent with strong trends of the form $x_t = x_{t-1} + \beta + \epsilon_t$ as in (1.12) when $\beta > 0$. Figure 3.3 records plots of x_t against x_{t-1} for p_t, w_t, g_t, l_t, k_t and $(w-p)_t$.

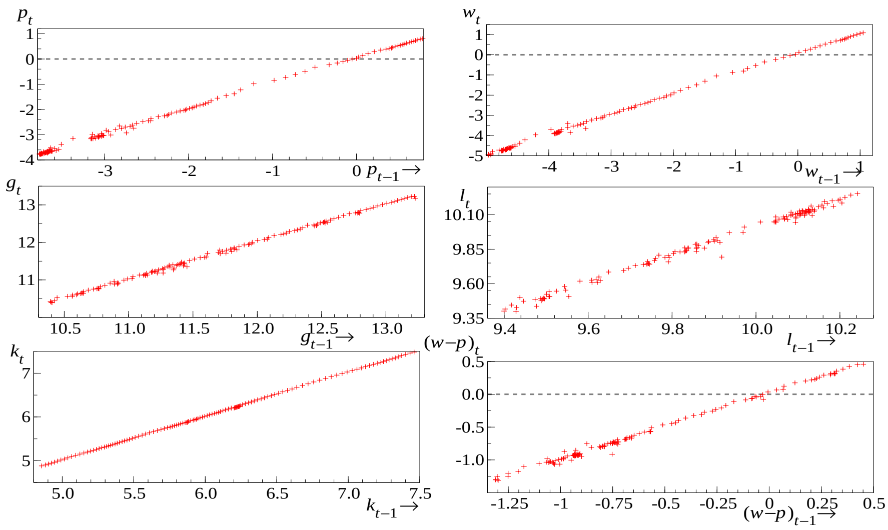


Figure 3.3
Plots of x_t against its lagged value x_{t-1}.

The graphs are almost perfect straight lines for all six variables plotted, despite their rather different time-series graphs seen above.

3.3 Correlograms

Once $\mathsf{corr}(x_t, x_{t-1})$ is thought about, one realizes that the string $\mathsf{corr}(x_t, x_{t-2})$, $\mathsf{corr}(x_t, x_{t-3})$, ..., $\mathsf{corr}(x_t, x_{t-s})$ say for any s, can be studied. When there are many lags, it is easiest to graph sample autocorrelations.

A plot of $\mathsf{corr}(x_t, x_{t-s})$ against s is called a correlogram, and Figure 3.4 shows some of these. The vertical axis plots $\mathsf{corr}(x_t, x_{t-s})$ and the horizontal axis shows the corresponding value of $s = 1, \ldots, 20$.[1] All the autocorrelations are close to unity for the whole lag length of 20 years, revealing almost complete 'persistence': $\mathsf{corr}(x_t, x_{t-20}) \simeq 1$ for these six series. The dashed lines show approximate 95% confidence bands for the null hypothesis that each autocorrelation is zero: all the autocorrelations lie far outside the null significance band throughout, and none 'die out' to zero as the lag length s increases. Why does that happen?

3.4 Explaining autocorrelations of levels

The population variance of ϵ_t, denoted $\mathsf{Var}[\epsilon_t]$ is defined for a stationary series by:

$$\mathsf{Var}[\epsilon_t] = \mathsf{E}\left[(\epsilon_t - \mathsf{E}[\epsilon])^2\right].$$

[1] The graphs are for correlations correctly defined, not the formulae often quoted in textbooks that rely on the time series being stationary–which these most definitely are not: see Hendry and Nielsen (2007).

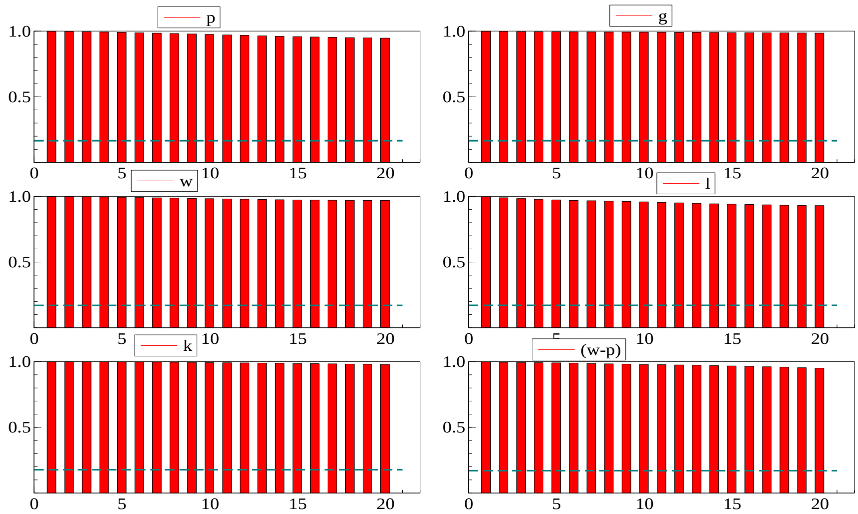


Figure 3.4
Correlograms of levels of variables.

However, when $x_t = x_{t-1} + \epsilon_t$, setting $\beta = 0$, we saw above that:

$$x_t = x_0 + \sum_{s=1}^{t} \epsilon_s \tag{3.2}$$

so that x_t depends on t epsilons and is not stationary. Nevertheless, as these $\{\epsilon_t\}$ are all independent and identically distributed, the variance of the sum is the sum of the variances, so that considering the 'population' here as $s = 1, \ldots, t$:

$$\mathsf{Var}[x_t] = t\mathsf{Var}[\epsilon_t].$$

There is a trend in the variance, which is why we say that x_t has a stochastic trend, despite no drift when $\beta = 0$.

Next, we also have that:

$$x_{t-20} = x_0 + \sum_{s=1}^{t-20} \epsilon_s \tag{3.3}$$

which has 20 epsilons fewer than (3.2), so that $\mathsf{Var}[x_{t-20}] = (t-20)\mathsf{Var}[\epsilon_t]$. Consequently, x_t and x_{t-20} share their first $(t-20)$ epsilons. Given a second stationary series u_t, the covariance between ϵ_t and u_t is:

$$\mathsf{Cov}(\epsilon_t, u_t) = \mathsf{E}\left[(\epsilon_t - \mathsf{E}[\epsilon_t])(u_t - \mathsf{E}[u_t])\right].$$

Applying this formula to the covariance between x_t and x_{t-20}:

$$\mathsf{Cov}[x_t, x_{t-20}] = (t-20)\mathsf{Var}[\epsilon_t].$$

Thus, all three components of the population correlation $\mathsf{Corr}[x_t, x_{t-20}]$ share $\mathsf{Var}[\epsilon_t]$, which cancels, so for the population (where sample averages are replaced by expectations over $s = 1, \ldots, t$):

$$\mathsf{Corr}[x_t, x_{t-20}] = \frac{\mathsf{Cov}[x_t, x_{t-20}]}{\sqrt{\mathsf{Var}[x_t]\mathsf{Var}[x_{t-20}]}} = \frac{(t-20)}{\sqrt{t(t-20)}} = \sqrt{1 - 20/t} \tag{3.4}$$

which is approximately unity when t is large. Consequently, stochastic trends can generate very high autocorrelations when t is large.

3.5 Correlograms of changes in the variables

Equally, differencing should remove the common cumulative $\{\epsilon_t\}$ (and the trend if $\beta \neq 0$) as:

$$\Delta x_t = x_t - x_{t-1} = \epsilon_t \tag{3.5}$$

so correlograms of changes, namely $\mathsf{Corr}(\Delta x_t, \Delta x_{t-s})$ should be small as there are no common cumulated shocks.

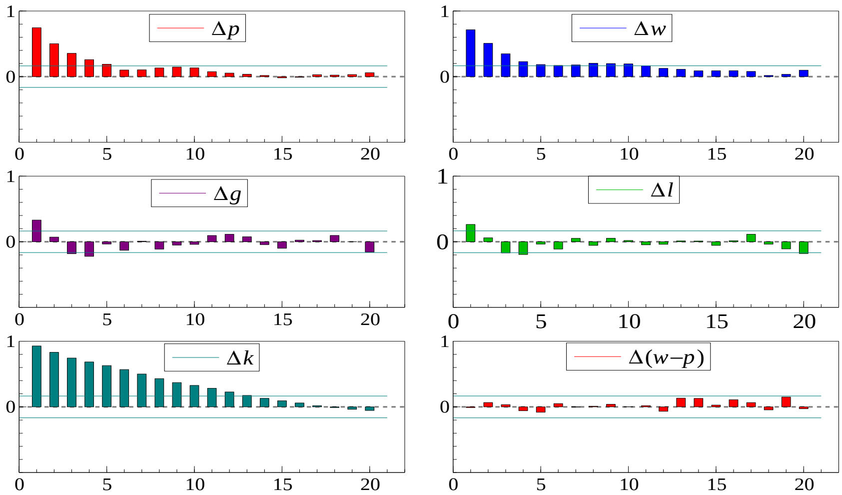


Figure 3.5
Correlograms of differences of variables.

Figure 3.5 shows the sample correlograms for the differences of the six main time series. These are very different from the autocorrelations of the levels, and now also rather different from each other. Wage and price inflation autocorrelations start high, and are positive for about 10 years or more—so there is some evidence of 'persistence'—whereas those for Δg_t, Δl_t are both near zero after one period. The correlogram for Δk_t (net investment) is more like that of an I(1) variable. Here all the autocorrelations 'die out' somewhat as the lag length s increases.

There is considerable 'cancellation' of the autocorrelations for $\Delta(w_t - p_t)$ relative to Δp_t and Δw_t. In our more technical terminology, therefore, $\Delta(w - p)_t$ has removed the common stochastic trends of p_t and w_t, most of their breaks, and their high autocorrelations.

3.6 Correlations between differenced variables

Correlations between variables are also greatly altered by differencing. Figure 3.6 records the scatter plots between $(\Delta x_{j,t}, \Delta x_{k,t})$, for $j \neq k$ where we have added

$\Delta U_{r,t}$, the change in the unemployment rate, which is the next variable we will consider. These plots are also completely different from the corresponding scatter plots between the levels, and again somewhat different from each other. The first three, for Δw_t & Δp_t, Δg_t & Δl_t, and $\Delta(w-p)_t$ & Δw_t are quite high with positive regression lines. The next three, between $\Delta U_{r,t}$ & Δl_t, Δg_t & $\Delta U_{r,t}$, and Δw_t & $\Delta U_{r,t}$, are also quite similar but now are negatively correlated. The final three, between $\Delta(w-p)_t$ & $\Delta U_{r,t}$, Δl_t & Δk_t, and Δk_t & $\Delta U_{r,t}$ are all relatively uncorrelated.

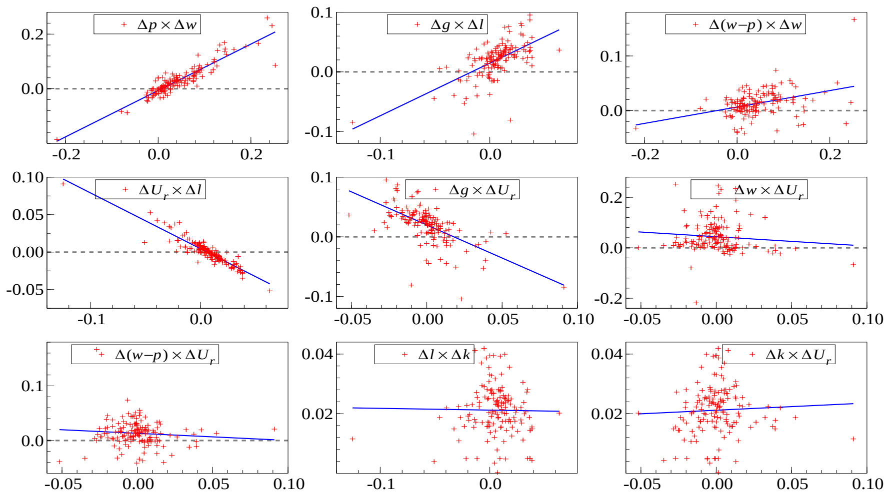


Figure 3.6
Many correlations between differences.

We will now apply what we have learned about analyzing time series to an initial appraisal of the time-series evidence on UK unemployment.

3.7 UK unemployment

Unemployment, $U_{n,t}$, is measured by the difference between the supply of labour willing to work, denoted L_t^s, and the demand for that labour, L_t^d, so $U_{n,t} = L_t^s - L_t^d$. The unemployment rate, $U_{r,t}$, is defined as $U_{n,t}/L_t^s$. Employment L_t is a derived outcome from the demand to produce goods and services, so varies with the profitability and volume of sales changes, as well as the technology embodied in the available capital stock and the skills of the available labour supply. It is often assumed that actual employment, L_t, equals L_t^d as firms do not employ unnecessary labour, but there is good evidence of cyclical 'labour hoarding'.

Total labour supply depends on the population of 'working age'—which has changed greatly over the last 150 years, when young children were often made to work in mines and factories—and on the 'willingness to work' at current wages. L_t^s is often measured by surveys of 'willingness to work', but $U_{n,t}$ is usually directly measured by registrations at unemployment offices (rather than indirectly as $L_t^s - L_t^d$), although earlier in our sample, unemployed numbers were based only on Trades Unions' records. More recent calculations of 19th Century unemployment in Boyer and Hatton (2002) show a very similar pattern with less volatility, as would be expected when unemployment is measured more generally. All historical data are estimates, prone to revision, but as Figure 3.7 shows, unemployment has changed so much over the long period we are analyzing that a fairly clear picture emerges.

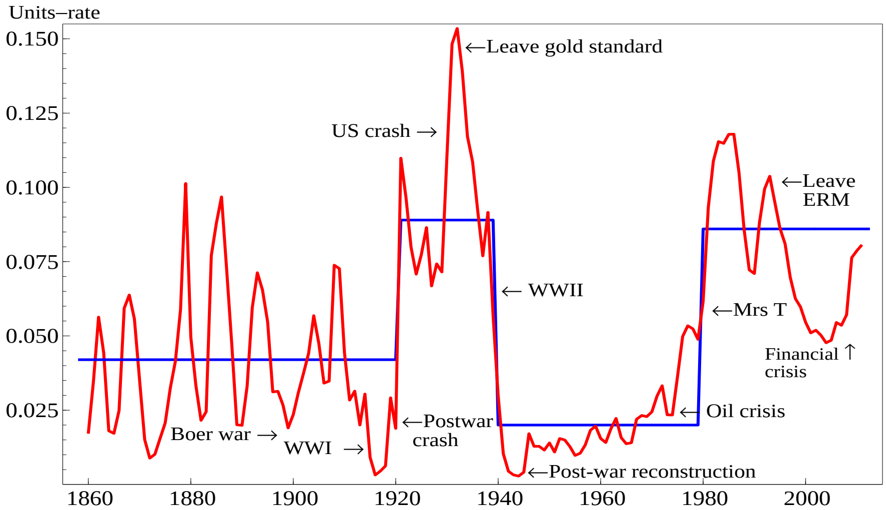


Figure 3.7
UK unemployment, 1860–2011, with major events affecting it.

As with earlier graphs, Figure 3.7 highlights a number of distinct regimes. The first, pre-WWI, reveals business-cycle behaviour of fluctuating levels of unemployment around a relatively constant mean of about 4.5%, with a sharp fall during WWI to near zero. Then there is a dramatic rise to more than 10% in the post-WWI crash, followed by a further increase during the Great Depression to a high of 15% in 1932. Unemployment then declined to high single digits after the UK abandoned the Gold Standard. The Second World War led to another dramatic drop to near zero, with unprecedently low levels persisting into the 1960s, accompanied by a much smaller variance than previously, under 'Keynesian policies'. The Oil Crises and Mrs Thatcher's economic policies combined to reproduce inter-war levels of unemployment, only slowly reduced after the UK left the Exchange Rate Mechanism (ERM) in 1992, before rising again during the recession that followed the 2008 financial crisis.

The shifts between the various regimes are abrupt, and they then persist, often for decades, so are most unlikely to be due to changes in 'willingness to work'. Hours of work generally fell, and paid holidays and the working age rose, reducing labour supply, offset by population growth and increasing female participation with increases in education and skills raising 'human capital'. Certainly, unemployment benefits increased as Figure 1.10 showed, but the common feature of the four main shifts are changes in demand for labour, driven by changes in the demands for 'goods and services' (if bullets, bombs, shells and soldiers could be so designated during wars).

There are in fact two key issues: what determines unemployment; and what does it determine? We will start with the second.

3.8 Macroeconomic consequences of unemployment

The most famous theory is that unemployment determines inflation, encapsulated in the 'Phillips' curve first proposed by Bill Phillips (1958) and often written as:

$$\Delta w_t = f(U_{r,t}) + e_t \tag{3.6}$$

where $f(\cdot)$ represents a non-linear relation with an ever increasing impact as unemployment falls towards zero. Figure 3.8 illustrates the relationship between nominal wage changes and unemployment levels over Phillips's sample 1860–1913, with a curve that represents the general tendency.[2]

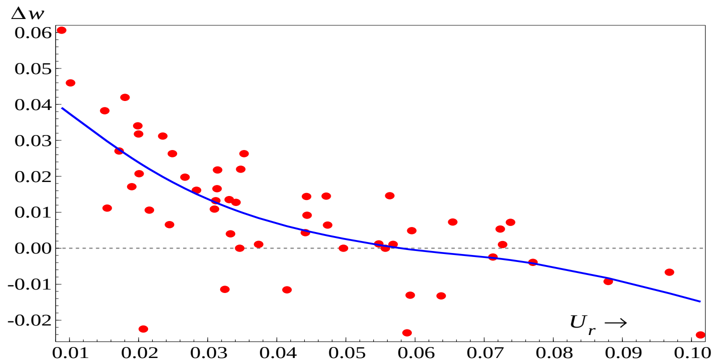


Figure 3.8
Replicating the original Phillips curve, 1860–1913.

The graph shows that lower unemployment, or higher demand for labour, leads to higher wage inflation. Phillips (1958) postulated a non-linear relation $f(\cdot)$ such that very low levels of **involuntary** unemployment had a more than proportional effect on nominal wage inflation, whereas very high levels of involuntary unemployment would have little impact on further reducing inflation. This relation has been seen as a 'trade-off' between inflation and unemployment: fewer unemployed entail more inflation. That hypothesis requires that the link is a causal one, and is constant over time. Both aspects are doubtful, as we now investigate.

3.9 Shifts in the inflation-unemployment relation

The Phillips curve seems fine for describing the late 19th and early 20th century. Before we consider the later evidence, remember that there were eight shifts in Δw_t but only four in $U_{r,t}$, so some shifts are not going to match very well, leading to what is sometimes called a 'breakdown' in the Phillips curve.

[2]For the technically minded, this is a cubic spline, not too different from the curve that Phillips originally fitted by an altogether different approach.

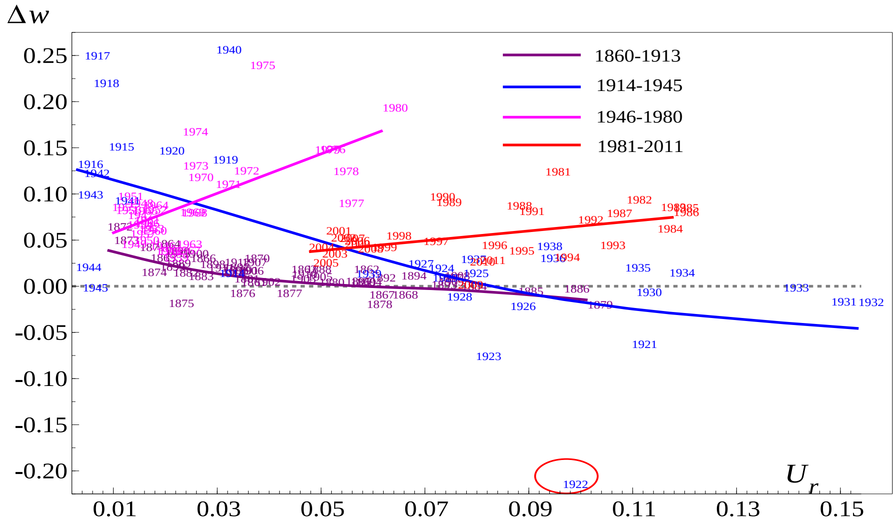


Figure 3.9
Shifts in inflation-unemployment 'trade-off'.

Figure 3.9 shows the curves for the four sub-periods of different unemployment behaviour, with the points shown by their dates. Two of the curves reveal an implausible 'rising' relation—more unemployment raises wage inflation. This could reflect that other factors affect their link—or that there is no genuine link. We also saw earlier that $\Delta(w - p)$ had only one major shift and a large spike, as against eight regimes for Δw, so price inflation must have had a major influence on nominal wage changes, a feature that is not included in the simple form in (3.6).

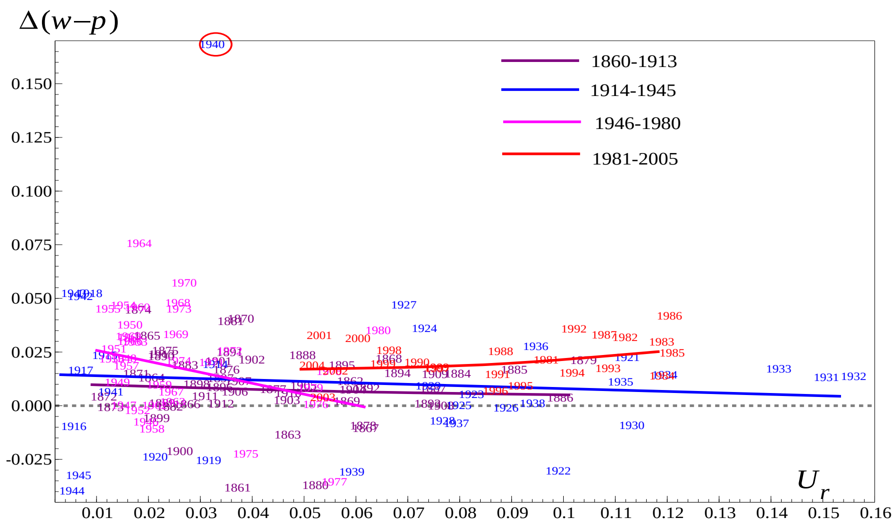


Figure 3.10
Shifts in the real-wage change-unemployment relation.

The notion of a trade-off between unemployment and real wages does not work either. Figure 3.10 shows the curves linking unemployment and real-wage changes for the same four sub-periods. It is slightly unclear what is happening because of the huge 1940 change, but there is little evidence of any relationship.

3.10 Other influences on the inflation-unemployment 'trade-off'

The graphs in Figure 3.10 also ignore our earlier finding that $w - p$ moves closely with $g - l$. Figure 3.11 records the link between $\Delta(w - p)$ and $\Delta(g - l)$, which is closer than that between real-wage changes and unemployment. Both regression lines of $\Delta(w - p)$ on $\Delta(g - l)$, and vice versa, are shown.

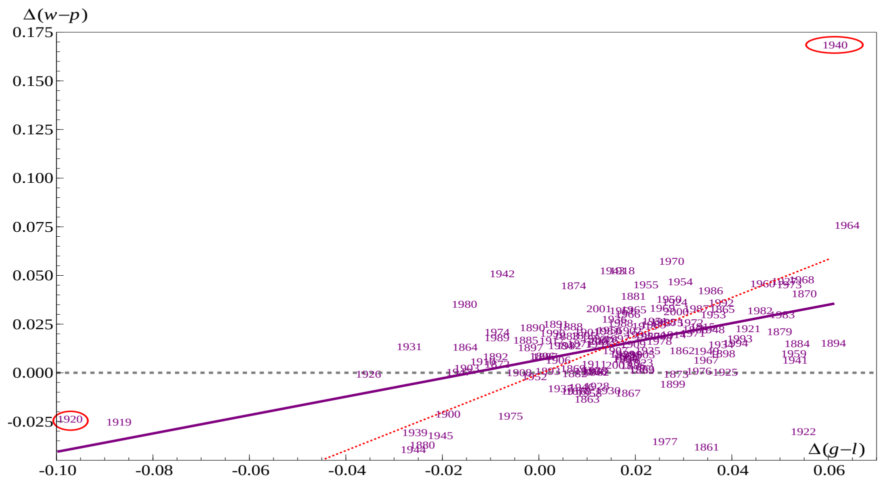


Figure 3.11
Changes in real wages versus changes in productivity.

The curves in Figure 3.10 also ignore any dynamic links, namely the close relations between any variable x_t and its lagged value x_{t-1}. Figure 3.12 joins each point to the next by decades over 1860–1913 to reveal 'Phillips loops'. The movements between the points are systematic, following a 'time line' that is shown by the dates, where each loops around its central curve.

The problem is that to understand how economics variables are really determined, an econometric relationship must allow for **all** the substantively relevant influences, any breaks, the dynamics of all inter-connections, any non-linear reactions, and any trends. The various simple descriptions above manifestly fail to do so. The next chapter explores how that can be done using modern approaches and software.

3.11 Chapter 3 key points

(A) There are high correlations between many macroeconomic variables, but these are not constant over time.
(B) There are even higher correlations between the levels of many economic variables and their lags: $\mathsf{corr}(x_t, x_{t-s})$ is often near unity despite $s = 20$ years.
(C) However, $\mathsf{corr}(\Delta x_t, \Delta x_{t-s})$ often 'dies out' rapidly as s increases.

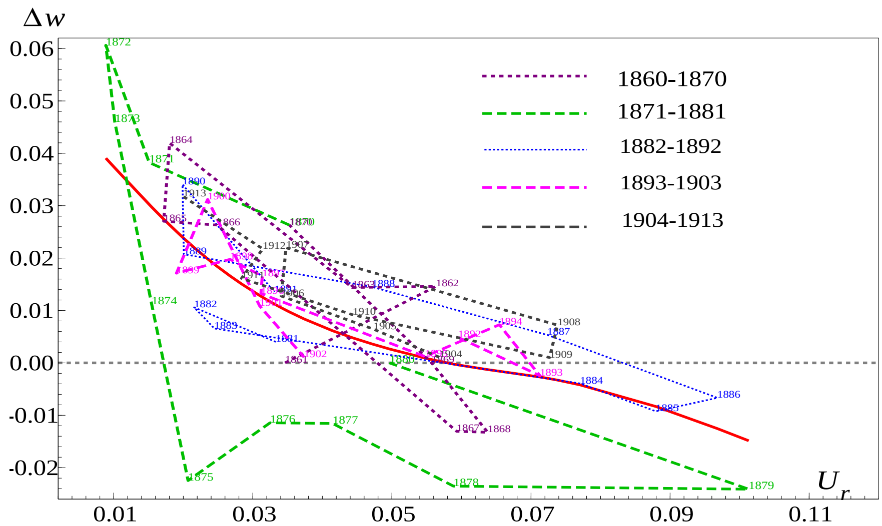


Figure 3.12
Dynamics of inflation-unemployment 'trade-offs'.

(D) Both (B) and (C) are consistent with macroeconomic variables having stochastic trends.
(E) Real-wage growth 'tames' the stochastic trends of w and p, most of their breaks, and even the autocorrelations of wage and price inflation.
(F) Conversely, both the correlations between, and autocorrelations of, differences of economic variables are often low.
(G) The unemployment-wage inflation relation known as the Phillips curve shifts over time.
(H) To understand the determinants of variables in economics requires models which include all substantively relevant variables, their dynamic reactions, all nonlinearities, breaks, and trends.

The next step, therefore, is to develop methods for handling these complications jointly, so in Chapter 4 we will consider the two main ingredients of macroeconometrics, namely economic theory and statistical analysis.

3.12 Task 9: Calculating data correlations

First, we will calculate the correlations in Table 3.1. Click on the `Model` icon, shown as a building block, fourth from the right on the icon line (or Alt+**Y**), to bring up the `PcGive - Models for time-series data` dialog (the default setting). All available modules are shown in its window by their icons (*PcGive* should be highlighted and named in the `Module` line). The `Category` should show `Models for time-series data`, and the `Model class` should be `Single-equation Dynamic Modelling using PcGive`. We want a different `Category`, called `Other models`, then the `Model class` should be changed to `Descriptive Statistics using PcGive`. Now select p, w, g, l, and k, click `OK`, then select `Means, standard deviations and correlations` (probably already marked), `OK`, and `OK` again to choose the full sample.

3.13 Task 10: Calculating sample correlograms

Next, we will replicate Figure 3.5 showing sample correlograms for the differences of the main time series. Click `Graphics`, select Δp, Δw, Δg, Δl, Δk, and Δwp, `All plot types`, `Time-series properties`, and tick `Create separate plots`, `Plot` and `Cancel`. Double click on the top-left graph to bring up `Edit Graph`, then `Copy properties to other areas`, select `World coordinates` and tick the '`Y`' box, then `Done`, so all the graphs are plotted over the range $[-1, 1]$.

3.14 Task 11: Replicating the original Phillips curve

We have learnt enough to reproduce the original Phillips curve shown in Figure 3.8. Click `Graphics`, select Dw and Ur, `All plot types`, `Scatter plots` and `Sample`, setting the dates to 1860–1913. Next, click `Smoothing`, and tick the `Cubic spline` box, then `Plot` and `Cancel`. Reset the lines, symbols, text and legends as desired.

To create Figure 3.9, first add *year* to the selection, `All plot types`, where `Scatter plots` appears, but now select `Two series by a third` and plot in turn for the four different sample sub-periods shown (remember to change the later date first), changing the graphics colours and line types in each so the lines can be distinguished. This will deliver a block of four graphs, then paste each on the first graph in turn (see Task 4 Section 1.19).[3] To circle an extreme outlier, such as 1922, right click for `Draw`, `Draw an Ellipse`, then start near to but below the date and move the pen up towards and past it: a little practice helps but incorrect attempts are easily deleted.

3.15 Task 12: Adding 'Phillips loops'

The final graphics sophistication is to reproduce Figure 3.12. Repeat the first part of Task 11 Section 3.14 for Dw and Ur over 1860–1913. Set the symbol size to 10, so the points are essentially invisible. Now add *year* and start the second part of Task 11, setting the sample sub-periods to 1860–1870, 1871–1881, ... ending with 1904–1913, so now you should have five graphs. For the four with dates, reset `Symbols` to `Line and symbols`, choose different colours for each, and set the symbol sizes for the dates to about 200. All the dates will now be joined up in order, so copy and paste each of the last four onto the first (Ctrl+**C**, Ctrl+**V** will speed the process). The Blue bent arrow on the Icon line can undo any mistakes. The `World coordinates` were enlarged in Figure 3.12 to clarify the loops.

[3] Figure 3.9 uses splines rather than regression lines, but there is not much difference here. To exactly replicate the figure involves first plotting each by scatter plots with splines, setting the symbol size to near zero so those points cannot be seen, then replotting each with dates, and pasting all onto the first graph.

3.16 Chapter 3 exercises

1. Do a scatter plot of nominal wage growth, 'Dw', against unemployment 'Ur'. Change the line `Type` from `symbols` to `line and symbols` and discuss what the time-linked lines show.
2. Fit a regression line with projections: does it change your interpretation in 1.?
3. Copy the graph in 1. and change to six regression lines with projections in the copy: can you explain why there is so little difference from 2.? Does it imply that there is a constant relation between Δw_t and $U_{r,t}$?
4. Return to the graph icon, and select `All plot types`>. Next, click on `Smoothing`, then tick `Cubic spline`, and `OK`. Is there much difference between the first linear regression in 2. and this graph?

Chapter 4
Two key ingredients: economic theory and statistical analysis

Chapter 4 guide posts

1. Economic-theory models of the behaviour of economic agents offer invaluable insights, but provide incomplete explanations for macroeconomic data: Section 4.1.
2. Statistical models are also theories, but of the data-generation process (DGP): Section 4.2.
3. Such models, like regressions equations, have unknown parameters that must be estimated from the available data: Sections 4.3–4.5.
4. The 'best' methods for doing so, and the resulting distributions of such estimators, are obtained assuming that the statistical model is the DGP: Section 4.6.
5. When the model does not represent the main features of the DGP, estimates can be poor and inferences about the model will usually be invalid: Section 4.7.
6. To ensure that the statistical model is a good approximation to the DGP, all substantively relevant variables, dynamics, breaks, non-linearities, and trends must be included: Section 4.8.
7. Statistical properties of estimators and tests can be hard to derive analytically, but are easily illustrated using simulation methods for computer-generated data, both for well-specified models (called congruent), and when important mis-specifications occur: Section 4.9.
8. Dynamic models are essential to characterize macroeconomic time series: Section 4.10.
9. Simple equations for unemployment demonstrate that not all models are equally useful, and that graphical statistics and formal mis-specification testing can reveal flaws in poor representations: Section 4.11.

4.1 Economic theory

Economic theory provides abstract models of the behaviour of economic agents, usually optimizing some objective. Such theories are often non-stochastic, assuming a 'steady state', and always have many (often implicit) *ceteris paribus* clauses. The aggregate is then taken to be a scaled up version of the micro-behaviour, although what may be optimal for one agent taking all others as given may be far from optimal if all try the same actions (as in many motorists simultaneously trying to avoid a traffic jam by using the same detour). Realistically, aggregation must be across heterogeneous individuals with conflicting objectives (as in buyers and sellers), whose endowments shift over time, often abruptly.

As a simple illustration, consider an economic analysis which suggests:

$$y = g(z) \tag{4.1}$$

where y depends on n 'explanatory' variables denoted z. The form of the function $g(\cdot)$ in (4.1) depends on the specific utility or loss functions of the agents, the constraints they face, and the information they possess. Analyses often just assume some specific form for $g(\cdot)$, that $g(\cdot)$ is constant, that only z matters in determining y, and that the z are 'exogenous', a term we discuss below, but taken to entail being determined outside the present analysis. Theories like (4.1) are a helpful starting point—but many aspects must be 'data determined', which requires postulating a statistical model of (4.1).

4.2 Statistical theory

A statistical model is also a theory—of how the data are generated. The simplest example for $y = g(z)$ in (4.1) is a linear model, with independent normal errors, a single exogenous regressor, z_t, and constant parameters, β_0 and β_1, such that:

$$y_t = \beta_0 + \beta_1 z_t + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right], \tag{4.2}$$

where successive ϵ_t are Independently drawn from a Normal distribution, with mean $\mathsf{E}[\epsilon_t] = 0$, variance $\mathsf{Var}[\epsilon_t] = \sigma_\epsilon^2$, and independent of z_t so that $\mathsf{E}[\epsilon_t z_t] = 0$.

The move from (4.1) to (4.2) also involved introducing the subscript t for time, so the relevant epoch over which (4.2) holds must also be specified, as say, $t = 1, 2, \ldots, T$ where 1 and T relate to precise historical time periods (usually 1860–2011 above). Moreover, the length of the time span from $t-1$ to t must be defined and often (as here with annual observations) depends on data availability rather than a 'natural' unit of time associated with agents' decision taking, although economic analyses rarely specify the precise meaning of t.

Although a formulation like (4.2) is simple, generalizations to many explanatory variables are easy, where z_t denotes a vector of variables as in (4.1). Already, (4.2) has some implications that need to be noted, the most important of which is seen on taking expectations of both sides conditional on z_t since (4.2) entails $\mathsf{E}[\epsilon_t | z_t] = 0$ leading to:

$$\mathsf{E}\left[y_t \mid z_t\right] = \beta_0 + \beta_1 z_t \tag{4.3}$$

as $\mathsf{E}[z_t | z_t] = z_t$. From (4.2), the variation around the regression line in (4.3) is normal and has a constant variance of σ_ϵ^2 (called homoskedastic, with the converse of a changing variance being called heteroskedastic).

The distributions of growth rates of aggregate data are often near normal after dealing with a few major location shifts. Figure 4.1 illustrates the data densities of the first and second differences of a number of UK macroeconomic variables with a reference normal density also shown as a dashed line. Then, using normality as a baseline, 'outliers' can be defined as substantial deviations (like the spike in real wages in 1940). Figure 4.1 shows that quite a few of the variables may be near to normal after removing a small number of outliers.

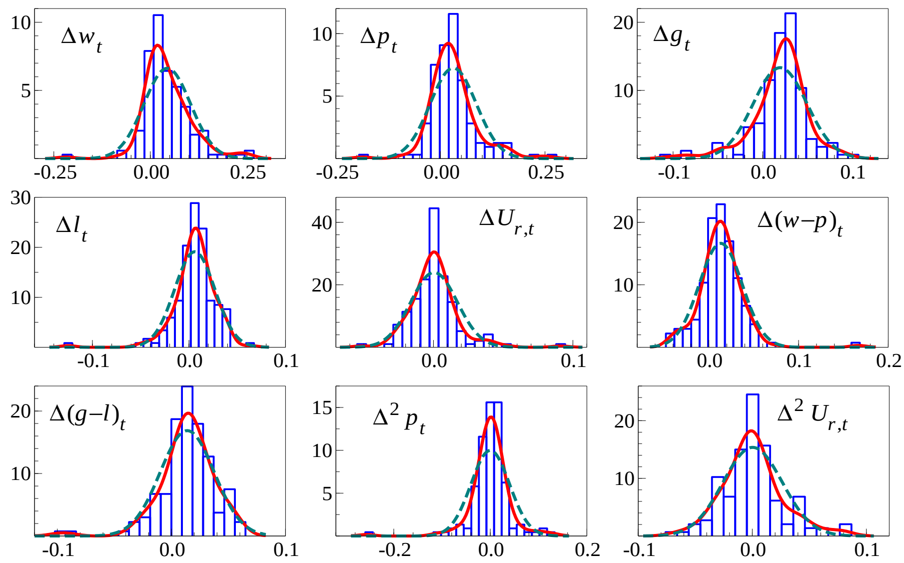


Figure 4.1
Densities of first and second differences of UK macroeconomic variables.

Importantly, the statistical properties of estimators of the parameters in models like (4.2) are well behaved for normal distributions, in which case (4.2) becomes a regression model, a case we now discuss.

4.3 The regression model

Conditional on z_t, taking the expectation of y_t in (4.2) yielded (4.3), so that:

$$y_t|z_t \sim \mathsf{IN}\left[\beta_0 + \beta_1 z_t, \sigma_\epsilon^2\right] \tag{4.4}$$

In (4.4), the statistical model entails that for a given value of z_t, observations on y_t are generated by a random 'shock' ϵ_t perturbing $\beta_0 + \beta_1 z_t$, possibly an agent's 'plan', to produce the independent outcomes $\{y_t\}$.

Even in this simple setting, there are three unknowns: the two regression parameters, β_0 and β_1, and the error variance, σ_ϵ^2. Systematic procedures for calculating the values of such unknown parameters from the data are now needed, assuming that the data are indeed generated by the statistical model. In turn, this requires well-defined criteria to determine which procedures are 'best'.

4.4 Estimation methods

An estimator is a function of the available data sample $(y_1 \ldots y_T, z_1 \ldots z_T)$—also denoted by $(\boldsymbol{y}, \boldsymbol{z})$ where the bold face denotes a vector—and the choice of that function determines the statistical properties of the resulting estimator, such as its mean and variance.

The three methods for estimating unknown parameters of statistical models that are the most common in econometrics are respectively called least squares, instrumental variables and maximum likelihood. We consider the first in section 4.5 and the second in section 4.8.3, but will not consider the third which is extensively discussed in Hendry and Nielsen (2007). Illustrations of least squares were provided by the trend lines fitted in Chapter 1, and by the regression line fitted to my name in Figure 1.4.

4.5 Least squares

In the simplest case where the value of β_0 in (4.2) is known to be zero (to simplify the algebra), the least-squares estimator (often denoted OLS, for ordinary least squares) for a sample of size T can be derived as follow. Multiply the regression equation:

$$y_t = \beta_1 z_t + \epsilon_t \tag{4.5}$$

by z_t to get:

$$z_t y_t = \beta_1 z_t^2 + z_t \epsilon_t \tag{4.6}$$

and average over the sample observations:

$$\frac{1}{T}\sum_{t=1}^{T} z_t y_t = \beta_1 \left(\frac{1}{T}\sum_{t=1}^{T} z_t^2\right) + \frac{1}{T}\sum_{t=1}^{T} z_t \epsilon_t \tag{4.7}$$

As $\mathsf{E}[\epsilon_t z_t] = 0$, so the population average covariance is zero, the estimator $\widehat{\beta}_1$ of β_1 is obtained by setting the corresponding sample average $1/T\sum_{t=1}^{T} z_t \epsilon_t$ to zero in (4.7) to obtain:

$$\frac{1}{T}\sum_{t=1}^{T} z_t y_t = \widehat{\beta}_1 \left(\frac{1}{T}\sum_{t=1}^{T} z_t^2\right) \tag{4.8}$$

and assuming $1/T\sum_{t=1}^{T} z_t^2 \neq 0$, solving for $\widehat{\beta}_1$ delivers:

$$\widehat{\beta}_1 = \left(\frac{1}{T}\sum_{t=1}^{T} z_t y_t\right) / \left(\frac{1}{T}\sum_{t=1}^{T} z_t^2\right) \tag{4.9}$$

To derive the properties of $\widehat{\beta}_1$ in (4.9) as an estimator for the unknown β_1, first substitute for y_t from (4.2), remembering that $\beta_0 = 0$, which delivers:

$$\widehat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T} z_t \epsilon_t}{\frac{1}{T}\sum_{t=1}^{T} z_t^2} = \left(\frac{1}{\frac{1}{T}\sum_{t=1}^{T} z_t^2}\right)\frac{1}{T}\sum_{t=1}^{T} z_t \epsilon_t \tag{4.10}$$

In (4.10), the $\{z_t\}$ are taken as fixed, as we are conditioning on them, so only the normally distributed $\{\epsilon_t\}$ are stochastic. Thus, $\widehat{\beta}_1 - \beta_1$ is a linear function of the

$\{\epsilon_t\}$ given in (4.2), from which we can obtain the properties of $\widehat{\beta}_1$ as:

$$\mathsf{E}\left[\widehat{\beta}_1\right] = \beta_1$$

so $\widehat{\beta}_1$ is an unbiased estimator of β_1 in this simple setting. Also:

$$\mathsf{Var}\left[\widehat{\beta}_1\right] = \frac{\sigma_\epsilon^2}{\sum_{t=1}^T z_t^2} \tag{4.11}$$

Thus, the distribution of $\widehat{\beta}_1$ is:

$$\widehat{\beta}_1 \sim \mathsf{N}\left[\beta_1, \frac{\sigma_\epsilon^2}{\sum_{t=1}^T z_t^2}\right] \tag{4.12}$$

and so:

$$\frac{\sqrt{\sum_{t=1}^T z_t^2}\left(\widehat{\beta}_1 - \beta_1\right)}{\sigma_\epsilon} \sim \mathsf{N}[0,1] \tag{4.13}$$

Although β_1 and σ_ϵ are never known, (4.13) points towards a useful way to test hypotheses about the value of β_1, such as the null hypothesis (denoted H_0) that H_0:$\beta_1 = 0$, as we will show shortly.

Next, the fit of the model is given by $\widehat{y}_t = \widehat{\beta}_1 z_t$. When the regression equation has an intercept like (4.2), the square of the correlation between y_t and $\widehat{y}_t$, denoted by R^2 and called the squared multiple correlation, is usually reported as a measure of the 'goodness of fit' of the model.

The residuals whose squares are minimized by least squares are given by:

$$\widehat{\epsilon}_t = y_t - \widehat{y}_t = y_t - \widehat{\beta}_1 z_t \tag{4.14}$$

and it can be proved that $\sum_{t=1}^T \widehat{\epsilon}_t^2$ is the smallest sum of squared residuals that can be achieved by the choice of $\widehat{\beta}_1$ in (4.9). Moreover we can estimate σ_ϵ^2 by:

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{(T-1)} \sum_{t=1}^T \widehat{\epsilon}_t^2 \tag{4.15}$$

The divisor of $(T-1)$ in (4.15) is called the degrees of freedom and is one less than T because we have estimated one parameter, $\widehat{\beta}_1$, which constrains the residuals.

The variance of $\widehat{\beta}_1$ in (4.11) can be estimated by:

$$\widehat{\mathsf{Var}}[\widehat{\beta}_1] = \left(\frac{\widehat{\sigma}_\epsilon^2}{\sum_{t=1}^T z_t^2}\right) \tag{4.16}$$

where the square root of the right-hand side is called the (estimated) standard error, ESE, of $\widehat{\beta}_1$ and denoted by $\mathsf{SE}[\widehat{\beta}_1]$ below.

Building on (4.13), using our estimate of σ_ϵ, and hence having an estimate of the denominator on the left-hand side, we can test the null hypothesis H_0:$\beta_1 = 0$ using:

$$\mathsf{t}_{\beta_1=0} = \frac{\widehat{\beta}_1}{\mathsf{SE}[\widehat{\beta}_1]} \tag{4.17}$$

which is distributed as Student's t under H_0 (rather than a standard Normal). Student was the pseudonym for W.S. Gosset, who first derived the distribution in (4.17), published as Student (1908). When the sample size T is large, Student's t is close to a standard Normal under the null, but allows for bigger departures from its mean of zero when T is small. The alternative hypothesis is $\mathsf{H}_1{:}\beta_1 \neq 0$ and when H_1 is true, (4.17) has a non-zero mean and the resulting distribution is said to be 'non-central'. The test procedure is to reject H_0 when the value calculated by (4.17) is far from zero. The probabilities of such departures from the mean under the null hypothesis are tabulated in many sources, and most econometric software packages calculate them automatically.

Two key values under H_0 for moderate T (larger than 50, say) are that $|\mathsf{t}_{\beta_1=0}| \geq 2$ occurs approximately 5% of the time, and $|\mathsf{t}_{\beta_1=0}| \geq 2.68$ about 1% of the time. As such a large value is likely to occur less than once in a hundred under H_0, it is more reasonable to assume the null hypothesis is false, so reject it in favour of the alternative that $\beta_1 \neq 0$.

We will not prove these results here, but illustrate them in the next section.

4.5.1 Interpreting regression estimates

What do such formulae entail when applied to actual data, and what properties do such estimators possess? It is easiest to understand estimator distributions from 'artificial data' of the kind we generated in section 2.9.1. Such data use random numbers for $\epsilon_t \sim \mathsf{IN}[0, \sigma^2_\epsilon]$, here setting $\sigma^2_\epsilon = 1$. To generate $\{z_t\}$, I used the equation $z_t = 0.005t + 0.8z_{t-1} + \nu_t$, where $\nu_t \sim \mathsf{IN}[0, 1]$. Then I set $\beta_0 = 1$ and $\beta_1 = 2$ to generate:

$$y_t = 1 + 2z_t + \epsilon_t \tag{4.18}$$

as in (4.2) when $T = 100$. This provides one trial from which we can estimate β_0 and β_1 by least squares. Here that outcome was:

$$\widehat{y}_t = \underset{(0.12)}{1.26} + \underset{(0.05)}{1.96}\, z_t$$

$$\mathsf{R}^2 = 0.942 \quad \widehat{\sigma}_\epsilon = 1.062 \tag{4.19}$$

In (4.19), $\widehat{\sigma}_\epsilon$ is the residual standard deviation, estimating σ_ϵ, and coefficient standard errors are shown in parentheses below coefficient estimates. The estimates of β_0, β_1 and σ_ϵ are quite close to the true values; and the fit is 'good' as measured by R^2.

4.5.2 Interpreting regression graphs

It is helpful to graph various aspects of that outcome. Figure 4.2 illustrates four features in its four panels.

Panel *a* records the outcomes y_t and fitted values $\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 z_t = 1.26 + 1.96 z_t$. As can be seen, the 'tracking' is close: $\widehat{y}_t$ is very similar to y_t.

Panel *b* shows the residuals $\widehat{\epsilon}_t = y_t - \widehat{\beta}_0 - \widehat{\beta}_1 z_t$ as in (4.14), but scaled by $\widehat{\sigma}_\epsilon$, so $\widehat{\epsilon}_t/\widehat{\sigma}_\epsilon$ are plotted. Consistent with panel *a*, the residuals look 'random' with most lying between ±2 (95% should do so if normally distributed).

Panel *c* shows the residual histogram and a smoothed density estimate, also showing a normal density for comparison.

Panel *d* reports the correlogram of the residuals: like panel *b*, there is little evidence of residual autocorrelation.

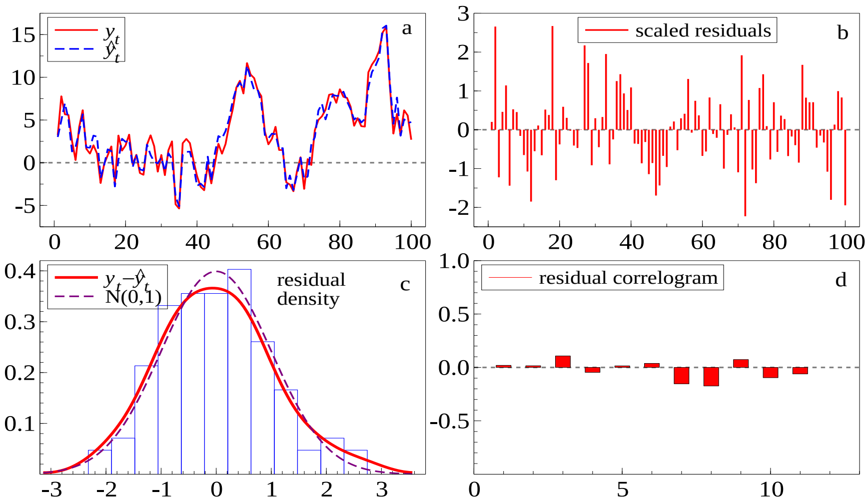


Figure 4.2
Fitted and actual values, residuals, residual density and correlogram.

All the graphs are consistent with the assumptions made, which is hardly a surprise as the model is a correct representation of the data generating process, and the data are accurately measured.

4.6 Econometric estimation in general

A model of $y|z$ is defined by its distribution $\mathsf{f}_{\mathsf{y}}(y|z,\beta)$, characterized by the parameters denoted by the vector β. In (4.4), $\mathsf{f}_{\mathsf{y}}(y|z,\beta)$ is $y_t|z_t \sim \mathsf{IN}\left[\beta_0+\beta_1 z_t, \sigma^2_\epsilon\right]$ for $t=1,\ldots,T$, where the unknown $\beta=(\beta_0,\beta_1)$, which needs to be estimated from the sample data $(y,z)=\left(y_1\ldots y_T, z_1\ldots z_T\right)$ by estimators $\widehat{\beta}=g(y|z)$.

Estimators and tests have sampling distributions such as $\mathsf{f}_\beta(\widehat{\beta}|\beta)$ for $\widehat{\beta}$ as in (4.13). Such distributions assume the model $\mathsf{f}_{\mathsf{y}}(y|z,\beta)$ is the process which generated $(y_1\ldots y_T)$, namely the data-generating process, denoted by $\mathsf{D}_{\mathsf{y}}(y_1\ldots y_T|z,\theta)$ where θ are the parameters of the economic agents' decision rules in that DGP. Estimators are then derived as if $\mathsf{f}_{\mathsf{y}}(y|z,\beta)$ was the DGP, which is what we did in Section 4.5.

'Good' estimators correctly reflect 'true' but unknown parameter values. For example, an unbiased estimator where $\mathsf{E}[\widehat{\beta}]=\beta$, ensures the average estimated value is the correct parameter. Another desirable property of an estimator is a small variance. However, in economics, interest really focuses on how close β is to the unknown θ, and how constant β is over time and across different policy regimes, issues we will address in the remainder of this chapter, particularly Section 4.9.

4.6.1 Distributions of regression estimates

When the data are artificial, we can generate thousands of samples of $\{y_t, z_t\}$, $t = 1, 2, \ldots, T$, estimate the parameters in every sample, and graph the resulting distributions in what is called a Monte Carlo simulation (yes, it is named after the famous Casino, as random numbers are used). Figure 4.3 illustrates the density of $\widehat{\beta}_0$ in panel *a* when estimating (4.18), showing that it is close to normal with a mean of 1 and a standard deviation of 0.14 (somewhat above $1/\sqrt{T} = 0.1$ as z_t does not have a mean of zero). Next, the density of $\widehat{\beta}_1$ is shown in panel *b*, and is again close to normal, but now with a mean of 2 and a standard deviation of 0.06. Third, the density of $\widehat{\sigma}_\epsilon$ appears in panel *c*, which is actually quite close to normal with a mean of 1 despite being an always-positive standard error. However, the density of R^2 in panel *d* is skewed left with a mean of 0.92.

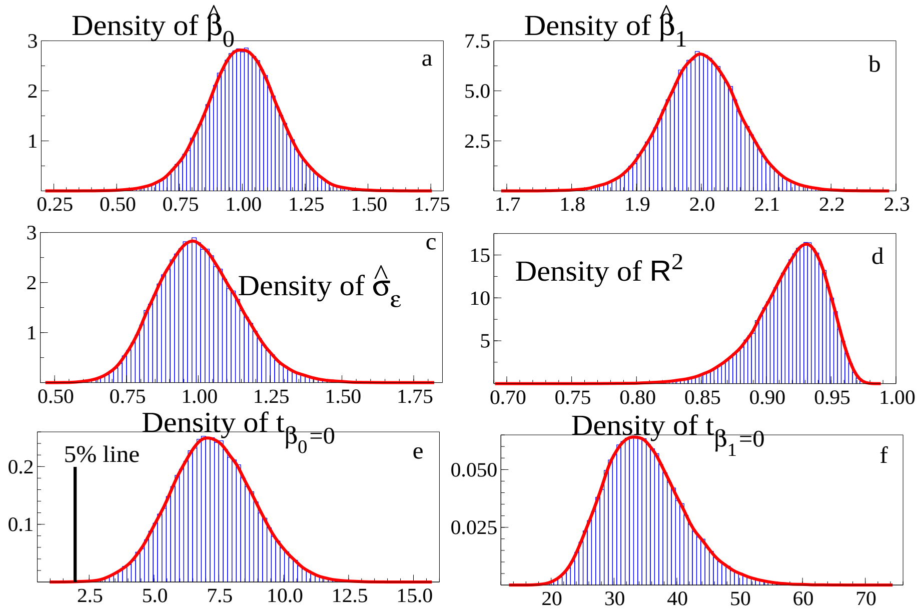


Figure 4.3
Distributions of parameter estimates and tests.

We considered tests of the form $\mathsf{t}_{\beta_1=0}$ in (4.17), and the densities for $\mathsf{t}_{\beta_i=0} = \widehat{\beta}_i/\mathsf{SE}[\widehat{\beta}_i]$ are shown in panels *e* and *f* respectively. These are distributed as Student's t under the null hypotheses that $\beta_i = 0$. That is false here for both parameters, and indeed the two tests reject their null hypotheses almost 100% of the time. Overall, our one sample in (4.19) was 'representative'.

4.7 Statistical models and data properties

Derivations like those in (4.5)–(4.15) require that the model is the DGP, so:

$$\mathsf{f}_\mathsf{y}(y|z, \beta) = \mathsf{D}_\mathsf{y}(y|z, \theta) \tag{4.20}$$

such that both the distribution $\mathsf{f}_\mathsf{y}(\cdot) = \mathsf{D}_\mathsf{y}(\cdot)$ and the parameters are the same so $\beta = \theta$. Statistical derivations will not in general deliver the correct results, namely that their implications match what actually happens, if $\mathsf{f}_\mathsf{y}(y|z,\beta) \neq \mathsf{D}_\mathsf{y}(y|z,\theta)$. Econometrics has developed many tests of whether a model matches the associated DGP. In practice, such tests often reject, after which it is unclear what an empirical investigator should do to 'fix' the problem. The only viable solution is to specify an initial statistical model that is general enough to ensure that it is not immediately rejected as failing to represent the DGP.

Since interdependence is a standard aspect of macroeconomic data, as our already examining a range of variables including $w,\ p,\ g,\ l,\ k,\ U_r,\ e,\ R_L, R_S,\ p_w$ reflects, statistical models often involve many variables and require multiple regression. We have also seen that time-series variables are highly dependent over time, that trends and breaks are common features, and that models need to be dynamic because reactions are not instantaneous, as seen in the Phillips loops: consequently statistical models must include all these features to match 'realistic' economic data.

4.8 Empirical econometrics confronts important difficulties

Among the most important difficulties are:

(A) Econometric models must allow jointly for all the complications of stochastic trends, breaks, dynamics, non-linearity, and interdependence.

(B) As the macroeconomy DGP is not known, investigators must specify a class of models general enough to include the DGP, then select the 'best choice' from that class.

(C) Economic theories are too abstract to adequately characterize the DGP. Fortunately, useful economic insights can be 'embedded' in the general model in (B) as part of the explanation, while letting the empirical evidence guide what additional aspects are needed.

(D) A general class of models may include dozens, perhaps hundreds, of variables to capture (A), far too large for humans to handle, though not too big for computers to tackle.

(E) Finally, it is important to check that the model really does describe the available data, leaving open how to proceed if it fails to do so.

We will illustrate these difficulties shortly by modelling the unemployment rate.

4.8.1 Simultaneity induced correlations

Consider two variables y_t and z_t with the joint distribution:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim \mathsf{IN}_2 \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right] \tag{4.21}$$

In (4.21), each of y_t and z_t has a normal distribution, with means μ_1 and μ_2 respectively and variances σ_{11} and σ_{22}, where the two variables are written in a column (as a vector) for convenience. The main new feature in (4.21) is the covariance σ_{12} between y_t and z_t.

Above we related the regression model (4.2) to the conditional expectation of y_t given z_t in (4.4) which for the joint distribution in (4.21) leads to:

$$y_t \mid z_t \sim \mathsf{N}\left[\mu_1 + \beta_{12}\left(z_t - \mu_2\right), \omega_{11}\right] \tag{4.22}$$

where $\beta_{12} = \sigma_{12}/\sigma_{22}$ and $\omega_{11} = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, which is the conditional variance. (4.22) can be expressed as the regression model:

$$y_t = \beta_{11} + \beta_{12} z_t + \nu_t \quad \text{where} \quad \nu_t \sim \mathsf{N}\left[0, \omega_{11}\right] \tag{4.23}$$

with $\beta_{11} = \mu_1 - \beta_{12}\mu_2$, then (4.23) can be estimated by least squares from the available data yielding $\widehat{\beta}_{11}$, $\widehat{\beta}_{12}$, $\widehat{\omega}_{11}$ and R^2.

It is also possible to formulate the 'reverse' regression:

$$z_t \mid y_t \sim \mathsf{N}\left[\mu_2 + \gamma_{21}\left(y_t - \mu_1\right), \omega_{22}\right]. \tag{4.24}$$

In (4.24), $\gamma_{21} = \sigma_{21}/\sigma_{11}$ with $\omega_{22} = \sigma_{22} - \sigma_{21}^2/\sigma_{11}$. In turn, (4.24) can be expressed as the model:

$$z_t = \gamma_{11} + \gamma_{21} y_t + w_t \quad \text{where} \quad w_t \sim \mathsf{N}\left[0, \omega_{22}\right] \tag{4.25}$$

yielding estimates $\widehat{\gamma}_{11}$, $\widehat{\gamma}_{21}$, $\widehat{\omega}_{22}$ and R^2. Notice that R^2 is the same in (4.23) and (4.25)–as an exercise, prove the useful result that $\widehat{\beta}_{12} \times \widehat{\gamma}_{21} = \mathsf{R}^2$.

Both (4.22) and (4.24) are equally 'good' regressions in this context, which raises the issue as to which to use, and how to make that choice? Fortunately, it is rare in macroeconomics to have so little structure, namely no trends, breaks, or other determinants of either variable. If either variable depended on an 'outside' influence, say w_t, then one can select which equation to model, as we now discuss.

4.8.2 Exogeneity

On the information given in the previous section, it is impossible to choose between the two directions of regression because y_t and z_t are simultaneously determined. For example least squares will estimate the coefficients β_{11} and β_{12} of (4.23), but as y_t and z_t are jointly determined those parameters may not correspond to the agents' decision parameters, so may not be useful, as the following example shows.

Consider an economy where the agents made plans to keep their 'permanent' consumption $\mathsf{E}[y_t] = \mu_1$ proportional to their 'permanent' income $\mathsf{E}[z_t] = \mu_2$ such that $\mu_1 = \kappa\mu_2$, so κ is of considerable importance to economic policy. An investigator estimates the relation in (4.23) as it entails:

$$\mathsf{E}[y_t] = \beta_{11} + \beta_{12}\mathsf{E}[z_t] \tag{4.26}$$

expecting $\widehat{\beta}_{11} \approx 0$, so interprets $\widehat{\beta}_{12}$ as $\widehat{\kappa}$. However, when $\mu_1 = \kappa\mu_2$ what (4.23) really entails is:

$$y_t = \left(\kappa - \frac{\sigma_{12}}{\sigma_{22}}\right)\mu_2 + \frac{\sigma_{12}}{\sigma_{22}} z_t + \nu_t \tag{4.27}$$

Consequently, $\widehat{\beta}_{12}$ estimates the ratio of the error covariance to the variance of z_t. For $\widehat{\beta}_{12}$ to be an estimate of κ requires that $\kappa = \sigma_{12}/\sigma_{22}$, so the means $\mu_1 = \kappa\mu_2$ are linked by the same parameter as the covariance: $\sigma_{12} = \kappa\sigma_{22}$. If so, then indeed $\beta_{11} = 0$ and $\beta_{12} = \kappa$. If that condition holds, z_t is said to be weakly exogenous for the parameters of (4.23) (see Engle, Hendry, and Richard, 1983).

The key idea behind weak exogeneity is that the parameters of the y_t and z_t processes are linked in (4.27) unless $\sigma_{12} = \kappa\sigma_{22}$, as μ_2 enters both otherwise. Thus, if μ_2 were to shift, as a consequence of a changed economic policy say, the regression relation between y_t and z_t would shift as well, making (4.27) useless. Conversely, when weak exogeneity holds, a regression of y_t on z_t would capture the constant parameter κ of interest. When weak exogeneity holds across one or more shifts in the parameters of conditioning variables, it is called super exogeneity: Section 6.8.1 provides an example of testing that hypothesis. As an exercise, prove that then y_t cannot also be weakly exogenous for the parameters of (4.26). Thus, one solution to the 'direction of regression' issue is to establish suitable weak exogeneity conditions.

The other solution is to determine an outside variable w_t that is weakly exogenous for the parameters of one of the y_t and z_t equations. If, say, w_t directly affected z_t, but was incorrectly excluded from (4.24), and only influenced y_t through z_t, then one could model the equation in (4.23) using that knowledge. Thus, when w_t is weakly exogenous for the parameters of (4.23), even though z_t is not, w_t can be used in an alternative estimation method known as instrumental variables.

4.8.3 Instrumental variables

We will again assume all means are zero to simplify the algebra. The regressor variable, z_t, may be correlated with the error on the equation in (4.22) re-written as:

$$y_t = \beta_1 z_t + \nu_t \tag{4.28}$$

because of their joint determination in (4.21), and hence $\mathsf{E}[z_t\nu_t] \neq 0$. Thus, least-squares estimates of β_1 will be biased. Instead of (4.6), instrumental variables uses another variable, w_t, where $\mathsf{E}[w_t\nu_t] = 0$, to multiply by:

$$w_t y_t = \beta_1 w_t z_t + w_t \nu_t \tag{4.29}$$

and average over the sample data:

$$\frac{1}{T}\sum_{t=1}^{T} w_t y_t = \beta_1 \left(\frac{1}{T}\sum_{t=1}^{T} w_t z_t\right) + \frac{1}{T}\sum_{t=1}^{T} w_t \nu_t \tag{4.30}$$

Assuming that $\mathsf{E}[w_t\nu_t] = 0$ is correct, the estimator is obtained by setting the corresponding sample average to zero:

$$\frac{1}{T}\sum_{t=1}^{T} w_t \nu_t = 0 \tag{4.31}$$

so that:

$$\widetilde{\beta}_1 = \frac{\sum_{t=1}^{T} w_t y_t}{\sum_{t=1}^{T} w_t z_t}. \tag{4.32}$$

A solution to (4.32) will exist only if $\sum_{t=1}^{T} w_t z_t \neq 0$, whereas $\sum_{t=1}^{T} z_t^2 \neq 0$ provided any $z_t \neq 0$. Thus, the two key requirements for a viable instrumental variable w_t is that it is correlated with the regressor z_t but not correlated with the error on the equation of interest. Remember that the very need for using an instrumental variable was that z_t was correlated with that error, so w_t has to steer a fine line of being correlated only with that component of z_t that is unrelated to the error.

Squared residuals are not minimized by instrumental variables, but residuals are still defined by:

$$\widetilde{\nu}_t = y_t - \widetilde{\beta}_1 z_t \tag{4.33}$$

and when there is only one regressor:

$$\widetilde{\sigma}^2_\nu = \frac{1}{(T-1)} \sum_{t=1}^{T} \widetilde{\nu}^2_t \tag{4.34}$$

The only equivalent of R^2 in equations with intercepts is the squared correlation between y_t and $\widetilde{y}_t$.

4.9 Congruent models and modelling

Given the turbulence that has afflicted the last 150 years as seen in graphs above, and the absence of evidence that unanticipated large shocks have ceased (*viz* the recent Financial Crisis), many features of macroeconomic data will not be included in economic analyses, however good such theories may be. Thus, econometric modelling is not simply the science of 'estimating a fully-specified theory-model' by choosing the 'best estimator', but requires tools to 'match' a model to the available data within the theoretical framework adopted.

The first step is to define what is meant by a 'match' when the theory-based statistical model specification does not coincide with the DGP. There is a large literature devoted to debating that issue, but here we will focus on what is known as the theory of reduction, that relates any postulated model to the DGP from which it was derived by a series of reductions: see e.g., Hendry (2009). As the DGP is very large and undoubtedly very complicated for non-stationary data, the reductions mainly concern eliminating variables that are not substantively relevant to the model under analysis. If that can be achieved, the resulting model will provide a good approximation to the DGP of the set of variables under consideration; otherwise not. Many aspects of model specifications are open to empirical evaluation.

A key aspect of such reductions is that they lead to six main characteristics that any empirical model should satisfy if it is to be a good representation of the DGP for the variables bring studied. We will first list then discuss these:

I homoskedastic, innovation errors $\{\epsilon_t\}$;

II exogenous conditioning variables z_t;

III constant, invariant parameters of interest, β;

IV theory-consistent, identified structures;

V data-admissible on accurate observations; and

VI able to explain findings of rival models (encompassing).

We now discuss these in turn.

Errors like $\{\epsilon_t\}$ are homoskedastic when they have the same variance at all points in the sample (here over time) so $\mathsf{E}[\epsilon_t^2]$ is constant. Those errors are an innovation process (called a martingale difference sequence) when $\mathsf{E}[\epsilon_t \epsilon_s] = 0\ \forall t \neq s$, so are not autocorrelated.

A conditioning variable like z_t in (4.5), or w_t in instrumental variables estimation, needs to be exogenous, as discussed in section 4.8.2, so is determined 'outside' of the model under analysis. Exogeneity is not an easy concept (see the various definitions evaluated by Engle, Hendry, and Richard, 1983, and the application to non-stationary processes in Hendry, 1995b), and weak exogeneity can be hard to test in some settings, but that the conditioning variable is independent of $\{\epsilon_t\}$ is sometimes sufficient.

A parameter β is constant if it does not change over the sample, and is invariant if β remains constant when the process determining z_t is changed. We saw above that the presence in a conditional model of the parameters from the process generating the conditioning variable will preclude invariance to location shifts, as will happen when z_t is a variable used for economic policy.

A model is said to be congruent if it satisfies these first three conditions. This uses the analogy to triangles, where two are congruent if they precisely match after suitable rotations: however, one triangle may actually be the side of a pyramid, so the match is only in two dimensions, and there is no match at all in the third. Similarly a congruent model matches the DGP in aspects that have been checked, but may not do so in directions that remain unknown, so cannot be said to be 'true' in any useful sense.

A structure is an invariant system, which is theory-consistent when it does not contradict the theory from which it was supposedly derived, and is identified when its parameters are a unique representation (so different parameters are not equally good).

A data-admissible model is one that cannot generate impossible values, such as negative fitted values for U_r, which is constructed to be positive.

Finally, there are almost always competing explanations of macroeconomic phenomena. If one knew the DGP, it would be straightforward to account for all the existing empirical findings. A more demanding requirement, called encompassing, is to ask if one model can account for the findings of all rival models, which would be feasible if that model was indeed the DGP.

4.9.1 Evaluating empirical models

Econometric modelling involves discovering sustainable relationships, and rejecting models otherwise—so both construction and destruction are needed. We consider destruction first: while it will not make you friends, it may influence people, and is far easier than constructing useful empirical models. Evaluation tests if any proposed empirical model is congruent, so 'matches' the evidence. Corresponding to I–VI in the previous section, there are six null hypotheses to test:

i) the past: fit should only deviate from observations by homoskedastic, innovation residuals;
ii) the present: contemporaneous variables should be exogenous;
iii) the future: parameters should be constant and invariant to policy interventions;
iv) theory information: models must have unique parameters related to theory;
v) measurement information: observations should be accurate, and the model 'data admissible';
vi) encompassing: the specified model should account for the findings of other models.

Non-congruent models are ones that fail the evidence on *i*)–*iii*): see (4.47) below for an example. We can now consider the more positive process of how to construct a model that might survive the tests in *i*)–*vi*).

4.9.2 General-to-specific modelling

General-to-specific, *Gets*, is an approach to constructing congruent relationships that begins from the most general unrestricted model (GUM) that it is reasonable to postulate initially, given the sample size of data, previous empirical evidence, available economic theory, institutional knowledge and measurement information. A GUM tries to allow for everything that might matter substantively in a relationship, so has many variables, includes dynamics, breaks, non-linearities and trends, etc., but is formulated after a careful theoretical analysis and sensible data transformations. The aim of the GUM is to be sufficiently general that the DGP will be well approximated by it, preferably as a special case. Then reductions applied to simplify the GUM will eliminate unnecessary features to provide a parsimonious congruent representation of that DGP for the relevant variables.

To model the evidence, one first checks that the GUM is congruent, using the criteria in I–III, then it is simplified to a congruent, parsimonious, and interpretable, econometric model. In particular, parsimonious encompassing is defined by the selected model being able to encompass the GUM from which it was derived. The final model choice is tested as a valid representation of the DGP by seeing how well it can encompass other rival models; and when there have been any shifts in the processes determining the unmodelled variables, by also testing that the selected model is invariant to those shifts.

4.9.3 Automatic *Gets* model building

When the GUM is very large, the simplification process may be infeasible for a human investigator to undertake manually. Consequently, automatic techniques, programmed to build models by a general-to-specific model selection strategy without human intervention, are invaluable. Since every reduction stage is well defined, a computer can be programmed do it: *Gets* is just a step up from a computer calculating regression estimates as in (4.8) rather than doing so by hand. In fact, as computers can handle very large initial general models, it is feasible to allow for more variables in the GUM at the start than the number of observations, a case that will arise in Section 5.5.

Instead of a simple regression as in (4.5), you now need to imagine dozens of variables in the GUM as in:

$$y_t = \sum_{i=1}^{N} \beta_i z_{i,t} + \epsilon_t \qquad (4.35)$$

where N might be large and some of the $z_{i,t}$ may be lags and perhaps non-linear functions of variables. Then the computer eliminates all of the $z_{j,t}$ with estimated coefficients that are insignificant on t-tests at the pre-assigned level like 1%, while ensuring that the resulting model remains congruent, and can still parsimoniously encompass the initial GUM in (4.35).

4.10 Formulating dynamic models

So far in this chapter, we have only considered static equations of the form:

$$y_t = \beta_0 + \beta_1 z_t + \epsilon_t \tag{4.36}$$

Section 3.2 introduced autocorrelations between successive values of a variable, and the simplest formulation is an autoregressive equation such as:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \nu_t \tag{4.37}$$

where $\nu_t \sim \mathsf{IN}[0, \sigma_\nu^2]$ and $|\gamma_1| \leq 1$. In the easiest case where $\gamma_0 = 0$:

$$\mathsf{E}[y_t y_{t-1}] = \gamma_1 \mathsf{E}[y_{t-1}^2] = \gamma_1 \sigma_y^2 \tag{4.38}$$

so there is a first-order autocorrelation of γ_1. As an exercise show that the second-order autocorrelation is γ_1^2, and so on at longer lags declining as γ_1^s for s lags. A stable dynamic equation requires $|\gamma_1| < 1$, which is also needed for (4.37) to represent a stationary process.

We can combine (4.36) and (4.37) in a more general dynamic model called an autoregressive distributed-lag (often denoted ADL):

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \epsilon_t \tag{4.39}$$

In (4.39), y_t responds to z_t, to its own previous value, y_{t-1} (the autoregressive part), and to the lag z_{t-1} (the distributed lag). To be a stable equation requires $|\beta_2| < 1$ (if $\beta_2 = 1$, the equation needs to be expressed in differences; and $|\beta_2| > 1$ is an explosive process, which can occur in hyperinflations). That relation is then perturbed by the random error $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ as in (4.2). Thus, (4.39) adds dynamics (y_{t-1}, z_{t-1}) to inter-dependence (z_t), which are two of the four key ingredients that need to be included in any model that hopes to characterize the data we have graphed above. When z_t has a stochastic trend, (4.39) will reflect that as well, with y_t and z_t being cointegrated when $|\beta_2| < 1$ and $\beta_1 + \beta_3 \neq 0$.

4.10.1 Interpreting dynamic equations

To interpret a dynamic equation like (4.39), we will first transform it. The first step is subtracting y_{t-1} from both sides to create the first difference on the left-hand side:

$$y_t - y_{t-1} = \beta_0 + \beta_1 z_t + (\beta_2 - 1)\, y_{t-1} + \beta_3 z_{t-1} + \epsilon_t \tag{4.40}$$

Next, subtract $\beta_1 z_{t-1}$ from $\beta_1 z_t$, to create a difference, and to keep the equation balanced, add $\beta_1 z_{t-1}$ to $\beta_3 z_{t-1}$:

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t - (1 - \beta_2)\, y_{t-1} + (\beta_1 + \beta_3)\, z_{t-1} + \epsilon_t \tag{4.41}$$

which reveals that β_1 is in fact the impact of Δz_t on Δy_t.

Now collect the terms in y_{t-1} and z_{t-1} when $|\beta_2| < 1$ as:

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t - (1 - \beta_2)\left(y_{t-1} - \kappa_1 z_{t-1}\right) + \epsilon_t \tag{4.42}$$

where:

$$\kappa_1 = (\beta_1 + \beta_3)/(1 - \beta_2) \tag{4.43}$$

Finally, it is convenient to collect the intercept with the last term as well, writing (4.42) as:

$$\Delta y_t = \beta_1 \Delta z_t - (1 - \beta_2)(y_{t-1} - \kappa_0 - \kappa_1 z_{t-1}) + \epsilon_t \tag{4.44}$$

where $\kappa_0 = \beta_0/(1 - \beta_2)$. This formulation will prove important in interpreting dynamic equations and is called an equilibrium-correction mechanism.

When change ceases in (4.44), so $\Delta y_t = \Delta z_t = 0$, with no shocks, so $\epsilon_t = 0$ both $\forall t$, then (4.44) becomes $y = \kappa_0 + \kappa_1 z$, which is the equilibrium and is why the model in (4.44) is called an equilibrium-correction mechanism (often abbreviated to EqCM). When growth occurs, Δy_t 'corrects' to the previous deviation $(y_{t-1} - \kappa_0 - \kappa_1 z_{t-1})$ from equilibrium, at a rate depending on the magnitude of $(1 - \beta_2) > 0$, as well as reacting to Δz_t and ϵ_t. The test that $-1 < \beta_2 < 1$ will transpire to check for cointegration when the data are integrated (see Task 30 in Chapter 8).

Initially, EqCM derivations assumed that $\Delta y_t = \Delta z_t = 0$ occurred in a static world so $y_t = y_{t-1} = y$ and $z_t = z_{t-1} = z$, but that is unrealistic in economics. Instead, we consider a constant growth rate, where $\Delta y_t = \Delta z_t = g \; \forall t$ and $(y_t - \kappa_0 - \kappa_1 z_t)$ is a cointegrating relation, so both y_t and z_t are $\mathsf{I}(1)$. Now, (4.44) incorporates three of our main features, and we will return to how breaks are handled shortly. First, however, we will apply some of these developments to two simple models of unemployment.

4.11 First empirical models of the UK unemployment rate

We will estimate versions of two hypothetical theory models of the UK unemployment rate $U_{r,t}$:
the first is that a high wage share causes higher unemployment because labour is 'too expensive';
the second is that high unemployment leads to high unemployment from 'deskilled workers' being unable to obtain jobs.

Formulate the first theory model as the linear regression:

$$U_{r,t} = \mu_0 + \beta_1(w_t - p_t - g_t + l_t - \mu_1) + \epsilon_t = \beta_0 + \beta_1(w_t - p_t - g_t + l_t) + \epsilon_t \tag{4.45}$$

where μ_0 is the 'equilibrium' level of unemployment when the wage share $(w_t - p_t - g_t + l_t)$ is at its mean μ_1.

The second is specified as the autoregression:

$$U_{r,t} = \mu_0 + \gamma_1(U_{r,t-1} - \mu_0) + \nu_t = \gamma_0 + \gamma_1 U_{r,t-1} + \nu_t \tag{4.46}$$

Both are 'straw' examples, essentially designed to illustrate how *not* to proceed.

4.11.1 Wage-share model of unemployment rate

Estimation of (4.45) from the sample of data from 1860 to 2004 yields:

$$\widehat{U}_{r,t} = \underset{(0.10)}{0.40} - \underset{(0.05)}{0.19}\,(w_t - p_t - g_t + l_t)$$

$$\mathsf{R}^2 = 0.078 \quad \widehat{\sigma}_\epsilon = 0.034 \tag{4.47}$$

As earlier, estimated coefficient standard errors are shown in parentheses below coefficient estimates. The reported estimates 'seem significant'—in that the $\mathsf{t}_{\beta_i=0}$

statistics would reject their null hypotheses—but we will question that implication shortly. If they were significant, then apparently a high wage share would lower unemployment, which is the 'wrong' sign. The fit is very poor: an $\mathsf{R}^2 = 0.078$ suggests that most of the movements in unemployment are not explained by (4.47). Figure 4.4 records a range of graphical statistics which reveal numerous problems, as the various panels clarify that (4.47) is not really explaining the unemployment rate.

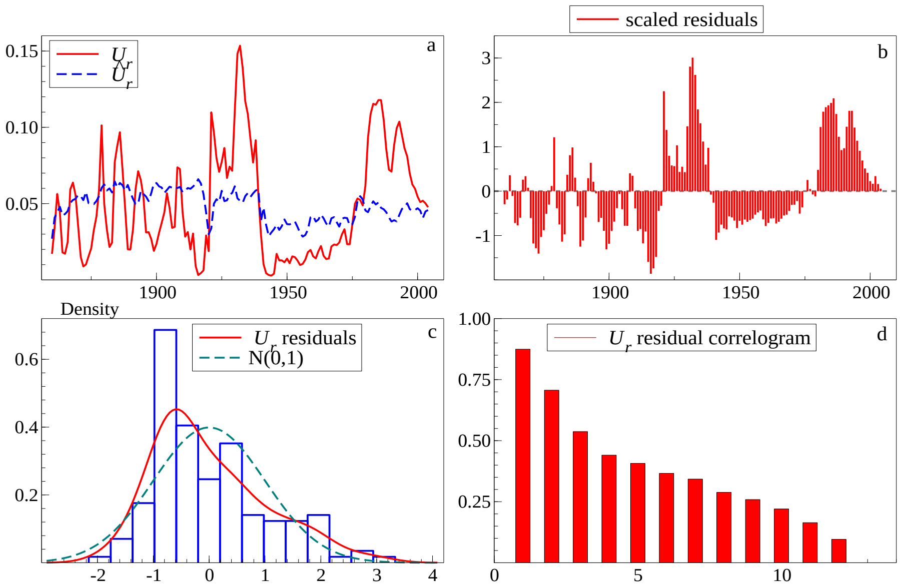


Figure 4.4
Wage share model of UK unemployment.

Panel *a* shows that there is almost no relationship between the fitted line, $\widehat{U}_{r,t}$, which is given by $0.4 - 0.19(w_t - p_t - g_t + l_t)$, and the data on $U_{r,t}$. In fact, $(w_t - p_t - g_t + l_t)$ does not have the correct 'time-series profile' to explain unemployment, as a comparison of Figure 2.7 with Figure 3.7 confirms. Thus, by itself $(w_t - p_t - g_t + l_t)$ cannot account for the behaviour of $U_{r,t}$, although that need not preclude it being a part of an explanation.

Because the fit is so poor, the scaled residuals, $(U_{r,t} - \widehat{U}_{r,t})/\widehat{\sigma}_\epsilon$, in panel *b* move systematically, and are far from 'random'. Panel *d* shows their correlogram is highly positively autocorrelated as far back as 10 years. Consequently, one of the crucial assumptions on which calculations of estimated coefficient standard errors depends is violated: the errors are not sequentially independent. Hence the so-called $\mathsf{t}_{\beta_i=0}$ statistics do not have their supposed Student-t distribution under the null, and different critical values would be needed for an appropriate test. The autocorrelation here is positive, and a more difficult exercise is for you to prove that then the estimated coefficient standard errors under-estimate the correct standard errors arising from sampling variability (see Tasks 16 Section 4.16 for simulations and 26 Section 7.10 for analysis).

Panel *c* plots the residual histogram, with an estimate of the density and a normal density for comparison. There is 'ocular' evidence of non-normality, so the explanation, such as it is, is uneven.

Conversely, unemployment by itself also cannot explain the wage share as the $R^2 = 0.078$ must be the same. Now, consider the performance of the ‘rival’ model in (4.46).

4.11.2 Autoregressive model of the unemployment rate

Estimation of (4.46) yields:

$$\widetilde{U}_{r,t} = \underset{(0.002)}{0.006} + \underset{(0.04)}{0.88}\, U_{r,t-1}$$

$$R^2 = 0.78 \quad \widehat{\sigma}_v = 0.016 \tag{4.48}$$

The fit is much better as measured by $R^2 = 0.78$ which is ten time larger, but more importantly, the residual standard deviation of 0.016 is less than half that of 0.034 for (4.47). Much of the movement over time in unemployment is explained by (4.48) as seen in Figure 4.5 panel *a*. The residuals in panel *b* are now less systematic, but there is a large ‘spike’ or ‘outlier’ in 1920, which remains even though least squares tries to minimize squared residuals, and hence ‘makes every effort’ to reduce the largest discrepancies. Consequently, nothing in the model can explain that jump in unemployment. The residual histogram in panel *c* is closer to the normal density, with a large outlier, and the residual correlogram in panel *d* is much ‘flatter’ than for (4.47). While far from a complete explanation of UK unemployment—unsurprisingly for such a simple model—it is obvious that (4.48) is much better than (4.47).

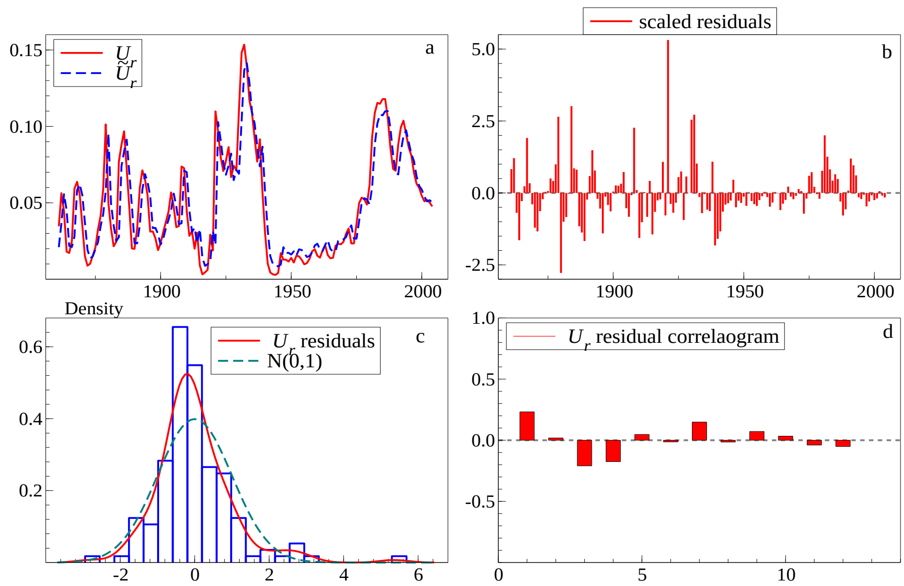


Figure 4.5
Autoregressive model of UK unemployment.

Here an encompassing test of whether (4.48) can account for the results in (4.47) can be conducted by including the wage share $(w-p-g+l)_t$ and its lagged value in the autoregressive model of $U_{r,t}$ in (4.48) and testing their significance. If either is significant, then the wage share adds to the explanation of unemployment, so (4.48) cannot mimic the DGP and explain even the poor fit of (4.47). Conversely, if neither matters, the wage share is irrelevant, and (4.48) 'predicts' that (4.47) will fit very poorly. Equation (4.49) records the outcome.

$$\begin{aligned}\widetilde{U}_{r,t} = &\underset{(0.04)}{0.89}\ U_{r,t-1} - \underset{(0.05)}{0.04} - \underset{(0.06)}{0.09}\ (w-p-g+l)_t \\ &+ \underset{(0.06)}{0.11}\ (w-p-g+l)_{t-1} \\ &\mathsf{R}^2 = 0.79\ \ \widehat{\sigma}_v = 0.016\end{aligned} \tag{4.49}$$

In fact, the wage share adds little to the fit of (4.48), with $\mathsf{R}^2 = 0.79$ when it was 0.78. Conversely, (4.47) cannot explain why (4.48) fits so well when it deems $U_{r,t-1}$ to be irrelevant: encompassing is asymmetric.

4.11.3 What lessons can we learn from these models?

First, and obviously, simple theories and models may not be sufficient to characterize macroeconomic data that are buffeted by many forces. Second, it is easy to see the failure of the simplest model in (4.47), from which we learn the important lesson that empirical findings can contradict the assumptions of the models on which they are based. Even the sign of the estimated coefficient of the wage share in (4.47) contradicted the theory. Third, some models describe the evidence much better than others, so we learn that not all models are born equal. Fourth, destructive testing can reveal the flaws in poor representations. However, an approach of postulating a sequence of simple models and finding they all fail is not a productive methodology, and could go on endlessly without delivering a useful outcome. Worse, it is difficult to judge the statistical properties of a later model, which may be designed to 'camouflage' a problem discovered in an earlier trial. Consequently, we need to allow for all the major determinants, stochastic trends, breaks, dynamics, non-linearities and interdependence jointly. While the prospect of such a large model may seem daunting, progress is feasible by letting the computer take the strain as we will see in the next Chapter.

4.12 Chapter 4 key points

(A) Economic-theory models can provide invaluable insights, but are only part of the explanation of macroeconomic data.
(B) Statistical models, $\mathsf{f}_{\mathsf{y}}(y|z,\beta)$, are theories of the data-generation process (DGP) $\mathsf{D}_{\mathsf{y}}(y|z,\theta)$.
(C) The unknown parameters β of such models need to be estimated from the available data.
(D) However, the 'best' methods for doing so, and the resulting estimator distributions, assume that the statistical model is the DGP.
(E) Statistical properties of estimators and tests can be easily illustrated using computer-generated data.
(F) To ensure that $\mathsf{f}_{\mathsf{y}}(\cdot) = \mathsf{D}_{\mathsf{y}}(\cdot)$, all substantively relevant variables, dynamics,

breaks, non-linearities, and trends must be modelled.
(G) The simple unemployment equations illustrated that not all models are equally useful, and that destructive testing can reveal the flaws in poor representations.
(H) Graphical statistics can reveal a great deal about how well an estimated model describes the associated time-series evidence.
The next step is apply these ideas to develop a more useful model of UK unemployment.

4.13 Task 13: Model formulation and estimation

We now move on to using *PcGive* for econometric modelling. Click on the 'building block', fourth from the right on the Icon line (or Alt+**Y**). The `Category` should show `Models for time-series data`, where the `Model class` should be `Single-equation Dynamic Modelling using PcGive`.[1]

Next, click on the `Formulate` button to bring up the dialog `Formulate - Single-equation Dynamic Modelling - UKHist2013.xls`. As with the `Graphics` dialog, `Selection` is the left column, `Database` the right, and the centre moves highlighted variables between them (or clears the selection). Unlike the `Graphics` dialog, variables are not temporarily removed from the Database, in case you want a longer lag (say). Set `Lags` to zero.

Double click on *Ur*, and then on *wpgl* to formulate (4.45) where we want to replicate (4.47): a Constant term is automatically added (double click on selected variables to remove them). `OK` brings up the dialog for `Model Settings`: for the moment just click `OK` again, to see the dialog for `Estimate` where you may need to set `Estimation ends at` to 2004, then `OK` to estimate the equation.

PcGive can record estimated equation output in several formats, including LaTeX, which can be included in TeX files to ensure fast and accurate reporting. Click on the `Test Menu` (second last on the right of the Icon line, or Alt+**T**) and tick the box for `Further Output...`. In that dialog, click on `Write model results` and select `LaTeX format`. Directly pasting that output into a TeX file as here would produce:

$$\mathrm{Ur} \;=\; \underset{(0.102)}{0.4048} \;-\; \underset{(0.0554)}{0.1926}\;\mathrm{wpgl}_t$$

The number of digits written to the output file can be set by changing the value opposite `Significant digits for parameters`, or `std.errors`. Other output and test outcomes will have to be added manually, but be careful with pasting statistics like Chiˆ2(2) to a TeX file without marking it as mathematics.

Next, `Part.Rˆ 2` is the squared correlation of each regressor with the dependent variable having removed the effects of the other regressors, so is the partial R^2. Also, `sigma` is the estimated equation standard error ($\widehat{\sigma}_\epsilon$), `RSS` is the residual sum of squares, the `F(1,150)` statistic tests the null hypothesis that all the non-constant regressors' coefficients are zero, `Adj.Rˆ 2` is R^2 adjusted for the number of estimated coefficients, `log-likelihood` is calculated assuming a normal distribution (see Hendry and Nielsen, 2007, for explanations and a likelihood-based approach), then there are notes of the numbers of observations and estimated parameters, and the mean and standard error of the dependent variable (here Ur).

[1] We now dispense with longer headers like '`Single-equation Dynamic Modelling`'.

To produce the estimation graphics, click on the `Test Menu` icon (second last on the right on the Icon line, or Alt+`T`), and tick the `Graphic Analysis...` box, then `OK` for the `Graphic Analysis` dialog. Two boxes for `Actual and fitted values` and `Residuals (scaled)` should be ticked: also tick the boxes for `Residual density and histogram (kernel estimate)` and `Residual autocorrelations (ACF)` to obtain the four-quadrant figure. The discussion in section 4.11 provided their interpretation. *PcGive* also reports more format tests of the assumptions in (4.2), so we consider these next.

4.14 Task 14: Model evaluation

This is a rather different Task, concerned with testing the specification of estimated models and interpreting the additional output produced beyond just coefficient estimates, their estimated standard errors and the ratios of those, denoted `t-value`.

Congruent models were discussed in section 4.9.1, which listed 6 null hypotheses to test. Here we consider *i*), that a model's fit should only deviate from the observations by homoskedastic, innovation residuals. The default setting for *PcGive* is to report various tests of *i*). The first is of innovation residuals against the alternative hypotheses that the residuals are autocorrelated, here a second-order autoregressive process reported as 'AR 1-2 test: F(2,148) = 267.55 [0.0000]**' (which we will write below as $\mathsf{F}_{\mathsf{ar}}(2, 148) = 267.6^{**}$) where the two asterisks denote that the null hypothesis of no autocorrelation is rejected at the 1% level, and the number in brackets '[0.0000]' is the probability of observing a value of 267.55 from a statistic with an F(2,148) distribution (see e.g., Godfrey, 1978, for details). This matches the manifest residual autocorrelation in Figure 4.4 panels *b* and *d*.

There are three tests of homoskedastic residuals against different heteroskedastic alternatives. The first is whether the squared residuals are autocorrelated, shown by 'ARCH 1-1 test: F(1,150) = 215.62 [0.0000]**', where ARCH is the acronym for Autoregressive Conditional Heteroskedasticity, and strongly rejects the null (see Engle, 1982), denoted $\mathsf{F}_{\mathsf{arch}}(1, 150) = 215.6^{**}$.

Before reporting the next two, there is a test of the residuals being normally distributed, shown as 'Normality test: Chiˆ 2(2) = 25.340 [0.0000]**', which again strongly rejects, consistent with Figure 4.4 panels *c* (see Doornik and Hansen, 2008), denoted $\chi^2_{\mathsf{nd}}(2) = 25.3^{**}$.

The two remaining heteroskedasticity statistics respectively test if the squared residuals are correlated with the squared regressors, or squares and cross products of regressors (which are the same here, but differ if there is more than one varying regressor) shown as 'Hetero test: F(2,149) = 1.5919 [0.2070]' and 'Hetero-X test: F(2,149) = 1.5919 [0.2070]' (see White, 1980), denoted $\mathsf{F}_{\mathsf{het}}(2, 149) = 1.59$.

Finally, there is a test for whether the relationship is linear based on testing the significance of adding the square and cube of the fitted value $\widehat{y}_t$ to the regression, shown as 'RESET23 test: F(2,148) = 14.453 [0.0000]**' which also rejects at 1% (see Ramsey, 1969), denoted $\mathsf{F}_{\mathsf{reset}}(2, 148) = 14.5^{**}$.

Rejecting on so many mis-specification tests raises an important methodological issue. Under the null that the model is correctly specified with homoskedastic, normally distributed innovation errors, then the tests of innovation errors, homoskedasticity and normality are nearly independent (i.e., in a simulation across many replications, there would be little correlation between the test outcomes).

However, once one of the assumptions is false, that no longer holds. For example, if there was an unmodelled location shift in the equation under test, so the intercept was not constant, then the residuals would be autocorrelated, non-normal, and heteroskedastic, and many of the above tests would reject. Consequently, it is incorrect to infer that findings for residuals entail that the errors have the same properties: autocorrelated and heteroskedastic residuals can occur with homoskedastic independent errors when there is an unmodelled location shift. 'Fixing' the problem of residual autocorrelation would not correct the more fundamental difficulty of the location shift. Thus, a strategy of generalizing simple models in response to failures of mis-specification tests does not have a sound basis (see e.g., Mizon, 1995).

To create and estimate the autoregressive model in (4.46), `Formulate`, and `Clear` the previous model selection if necessary, set `Lags` to one, and double click on *Ur*, then `OK`, `OK`, `OK` to replicate. Its 4-quadrant graphs follow as in Task 14 Section 4.14. Discuss its mis-specification test outcomes.

Finally, formulate and estimate equation (4.49) and its graphics.

4.15 Task 15: Monte Carlo simulations

We used artificial data in Section 4.6.1, and here we will apply the same Monte Carlo technique to simulate the distributions of the estimated parameters of a first-order autoregressive model (denoted AR(1)) like (4.37) mimicking (4.46) when the AR(1) is also the DGP:

$$y_t = \rho_0 + \rho_1 y_{t-1} + \epsilon_t \quad t = 1, \ldots, T \tag{4.50}$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ and $|\rho_1| \leq 1$. In the simpler case where $\rho_0 = 0$, the estimator of ρ_1 is based on:

$$\frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} = \rho_1 \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \right) + \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t$$

and since $\mathsf{E}[y_{t-1}\epsilon_t] = 0$, setting the last term to zero:

$$\widehat{\rho}_1 = \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} \Big/ \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \right) \tag{4.51}$$

A similar formula holds after taking deviations from means when an intercept is present as in (4.50).

To define the DGP in (4.50) for a simulation experiment requires specifying numerical values for ρ_0, ρ_1, T, and σ_ϵ^2, as well as the distributional properties of $\{\epsilon_t\}$ (here independent normal, with $\sigma_\epsilon^2 = 1$), and the initial condition y_0, which should be set at $\mathsf{E}[y] = \rho_0/(1-\rho_1)$ for a stationary process (here *PcNaive* discards the first $N = 20$ observations to offset starting from $y_0 = 0$). We will use $\rho_0 = 0$, $\rho_1 = 0.9$ with the largest $T = 50$.

Select `Model`, and change the `Category` to `Monte Carlo`, so the `Model class` should be `AR(1) Experiment using PcNaive`. The simulation program *PcNaive* is embedded within *PcGive*, but operates rather differently as we will see. Click `Formulate` which brings up the `Formulate - AR(1) Experiment` dialog. The top block concerns defining ρ_0 and ρ_1: the defaults are zero and 0.9 so we will first use those. The next block concerns the AR(1) model to be estimated: the defaults are including just y_{t-1} and no constant, which matches the DGP. The third block concerns the settings for the Monte Carlo experiment, namely the number of replications, and the value(s) for the sample size T (and how often any exogenous variable is to be renewed, which is not relevant here so shown in faint). A replication is a draw of T observations on (4.50) generating $y_1, \ldots, y_T$ from which one set of model parameter estimates can be calculated, denoted $\widehat{\rho}_{1,i}$. That is then repeated $i = 1, \ldots, M$ times, so set M to 1000. The line `Sample size: start:[step]end` allows many Monte Carlo experiments to be done at the same time across a range of sample sizes. Here, use `20:[1]50`, so the output will graph $\overline{\rho}_1 = \frac{1}{M}\sum_{i=1}^{M}\widehat{\rho}_{1,i}$ at $T = 20, T = 21, ...,$ up to $T = 50$.

Next, we set the `Monte Carlo Output`: tick the box for `Coefficients` if it is not already marked. Finally, tick the box for `Live Graphics` which will show the progress of the simulation as M increases across all the chosen values of the sample size. Tick the boxes for `Recursive coefficients and tests` (although we have not requested any tests yet), and `Histograms of coefficients and tests`, which shows the shapes of the densities of $\widehat{\rho}_1$ as T changes. Also tick the box for `Set plot frequency`, and set `Plot frequency` to 100, so every 100 replications the screen will show the latest updated results. Click `OK`, and for the simple experiment here, the simulation is run immediately and should only take a few moments on a modern PC. All the output is written to the Results file, but we will focus on the graphs here.

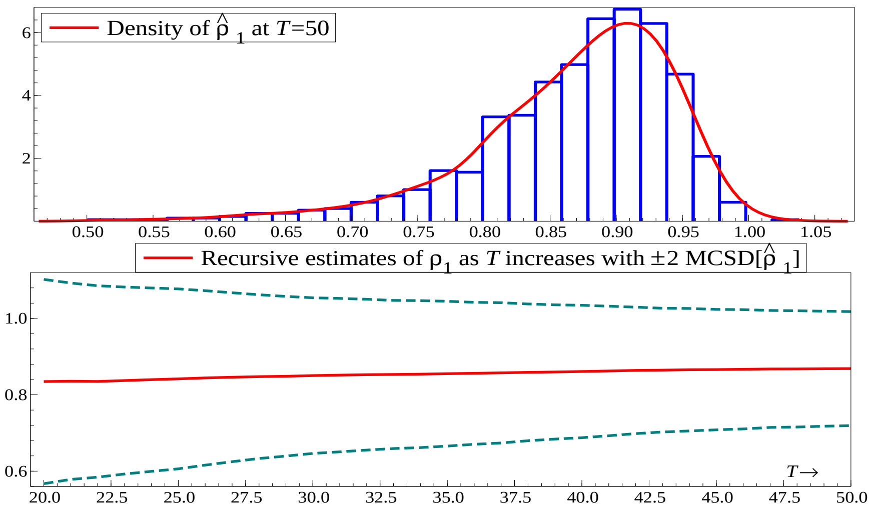


Figure 4.6
Simulation density and recursive estimates of $\rho_1 = 0.9$.

Two graphs should show on the screen, the top one for the density of $\widehat{\rho}_1$ as T changes, illustrating how that density shifts across sample sizes, and the lower for $\overline{\rho}_1$ at each sample size, which provides an estimate of $\mathsf{E}[\widehat{\rho}_1]$ at that sample size. Also shown on the graph are dashed lines at $\pm 2\mathsf{MCSD}[\widehat{\rho}_1]$ calculated from the variation across the M values of $\{\widehat{\rho}_{1,i}\}$ at each sample size (called MCSD, for

Monte Carlo Standard Deviation):

$$\mathsf{MCSD}[\widehat{\rho}_1] = \sqrt{\frac{1}{(M-1)}\sum_{i=1}^{M}\left(\widehat{\rho}_{1,i} - \overline{\rho}_1\right)^2} \tag{4.52}$$

At the end of the experiment, the top graph is the density of $\widehat{\rho}_1$ at T as shown in Figure 4.6 (edited for clarity). Even for $T = 50$, that density is quite skewed with a long left tail, with a few values outside the stationary range above unity. The lower graph shows that $\overline{\rho}_1$ started at about 0.81 at $T = 20$ and ended around 0.87 at $T = 50$, so throughout provided a downward biased estimate of the population parameter of $\rho_1 = 0.9$, but with the bias decreasing as T increased.

Now return to the `Formulate - AR(1) Experiment` dialog and this time include a `Constant` in the model, leaving the DGP value at zero (you may need to reset the `Plot frequency` at 100), and rerun. This time there are four graphs, densities for $\widehat{\rho}_1$ and $\widehat{\rho}_0$ and recursive graphs for each as shown in Figure 4.7 (unedited). The density for $\widehat{\rho}_0$ should be approximately normal, and its recursive mean is correctly close to zero throughout. It is the recursive graph for $\widehat{\rho}_1$ that has changed most: $\overline{\rho}_1$ now starts around 0.7 and ends at 0.82, so is much more biased. This downward bias is well known (see Hendry, 1984), and shows that the extent of autoregressive inertia is generally underestimated.

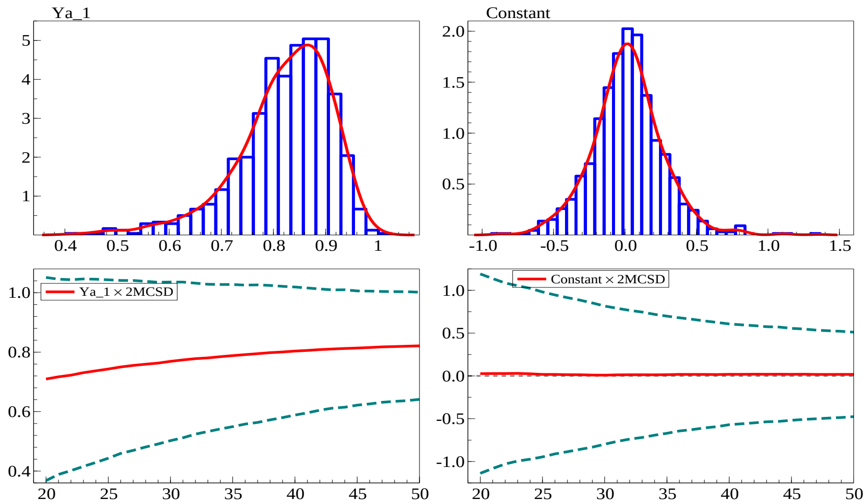


Figure 4.7
Simulation density and recursive estimates of ρ_1, ρ_0.

Monte Carlo extends the capabilities of econometricians to investigate problems that are analytically intractable, or extremely difficult. We can modify our present experiment a little to look at what happens to the density of $\widehat{\rho}_1$ when $\rho_1 = 1$ so the DGP is an integrated process. Select `Formulate - AR(1) Experiment` and change `Ya_1 coefficient` to unity in the block for `AR(1) DGP` leaving the `Constant` at zero, tick the box for `t-tests` and leave all other entries unchanged. Even though $\rho_1 = 1$, almost all estimates $\widehat{\rho}_1$ are less than unity, and the mode of the density is around 0.92. Any test for the presence of a unit root has to allow for that outcome. Further, the density of $\widehat{\rho}_0$ is non-normal, and has 'twin peaks' on either side of the true value of zero, especially marked on its 't-test' density, which rejects the null hypothesis much more often than a 5% significance level would suggest. A surprise is the 't-test' for $\mathsf{H}_0{:}\rho_1 = 0$ being skewed to the right.

The interesting test is of $\mathsf{H}_0{:}\rho_1 = 1$ which we cannot simulate on this simple design menu. The adventurous reader can do so by selecting `Model class` as `Advanced Experiment using PcNaive & Ox Professional`. *PcNaive* actually just designs Monte Carlo experiments, and writes a computer program in the language *Ox* (see Doornik, 2007, which is where the name and code for *OxMetrics* come from), then that program needs to be saved and executed by *Ox*, which requires *Ox Professional*. Task 16 in Section 4.16 will explain how to use this more flexible approach. However, even with the simpler dialog here, we can simulate one interesting property as follows. Return to set ρ_1 to 0.9, a model with a Constant, and set `Sample size:  start:[step]end` to 200 (so only the one value of T); run the experiment, change the colours of the density of $\widehat{\rho}_1$ by light shading inside the histogram using the `Edit Graph Menu`, and copy that density on top of the one for $T = 50$, then repeat all those steps again at $T = 800$ (now with dark shading), so the resulting graph shows the three sets of densities as the sample size has increased fourfold twice. Now repeat these two additional experiments with $\rho_1 = 1$ and collect all the densities of $\widehat{\rho}_1$ in one figure as shown in Figure 4.8.

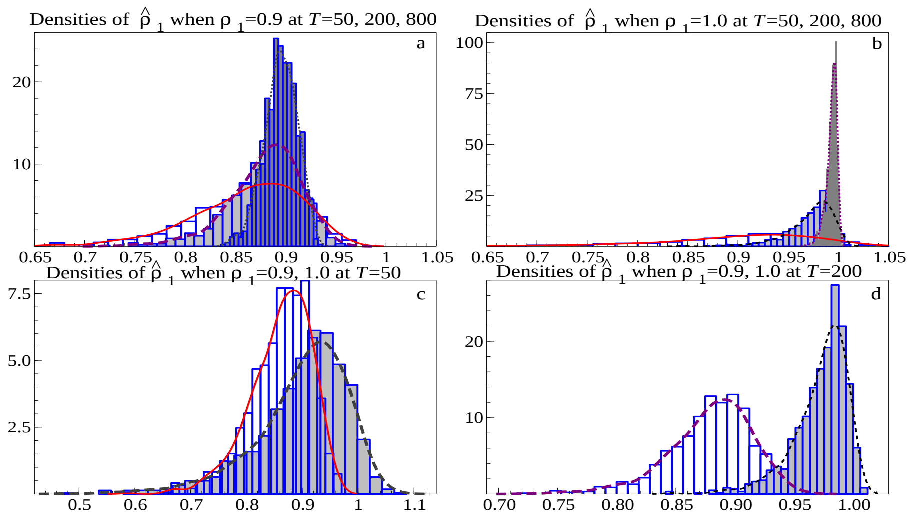


Figure 4.8
Simulation densities of $\widehat{\rho}_1$ when $\rho_1 = 0.9,\ 1$.

The markedly different shapes of the densities in the top row (panels a and b) are obvious, with the density when $\rho_1 = 1$ becoming much more concentrated in the neighborhood of unity as T increases than happens around 0.9 for $\rho_1 = 0.9$. Both graphs have been placed on the X-axis range 0.65–1.05 to highlight that property, known as 'super convergence', which arises from the integrated process cumulating all past shocks. That entails larger critical values are needed to correctly reject the null of a unit root in favour of a stationary process, or a cointegrating combination.

As the bottom row shows, at $T = 50$ there is considerable overlap between the densities of $\widehat{\rho}_1$ for $\rho_1 = 0.9$ (unshaded) and $\rho_1 = 1$, making discrimination between a stationary and an integrated process difficult, even at $T = 200$ (panel d).

4.16 Task 16*: Simulating the effect of error autocorrelation on ESEs

One of the crucial assumptions on which calculations of estimated coefficient standard errors depends is that the errors in a model are sequentially independent. If instead, errors are positively auto-correlated then the estimated coefficient standard errors (denoted ESEs) under-estimate the correct sampling standard errors arising from variability across samples. Consequently, the so-called $t_{\beta_i=0}$ statistics testing null hypotheses about parameter values do not have their assumed Student-t distribution under the null, and different critical values would be needed for an appropriate test. In this Task, we will check that claim by a Monte Carlo simulation, and Task 26 in Section 7.10 will analyze the downward bias in ESEs for the simplest univariate equation.

To perform this simulation, select `Monte Carlo` then the `Model class` as `Advanced Experiment using PcNaive & Ox Professional`. Task 15 Section 4.15 noted that *PcNaive* designs Monte Carlo experiments, then writes a computer program which needs to be saved and executed by *Ox*, so the present Task requires *Ox Professional* (the experimental design is stored in ESEBias.ox). The DGP is:

$$y_t = \beta x_t + u_t \tag{4.53}$$

where both $\{y_t\}$ and $\{x_t\}$ are stationary processes with:

$$u_t = \rho u_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{4.54}$$

such that $|\rho < 1|$ and:

$$x_t = \lambda x_{t-1} + \nu_t \text{ where } \nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right] \tag{4.55}$$

also with $|\lambda < 1|$, and $\mathsf{E}\left[\epsilon_t \nu_s\right] = 0 \; \forall t, s$. The selected DGP parameter values are $\beta = 1$, $\rho = 0.9$, $\lambda = 0.9$, with $\sigma_\epsilon = \sigma_\nu = 1$, $T = 20, \ldots, 100$ in blocks of 5 (so 20, 25, etc.), using $M = 1000$ replications.

The estimated model is (4.53) ignoring the error autocorrelation, simulated at a range of sample sizes to illustrate recursive Monte Carlo. In the `DGP Design` dialog, set `q` to 1, click on `PcNaive DGP Extras` and tick `ARMA errors`, then `OK`. In the next dialog for `DGP parameters`, set `A2` to 1 (corresponding to β), click `Y DGP errors` and set `B0` to 0.9 (corresponding to ρ), then click `Z DGP` and set `C0` to 0.9 (corresponding to λ), leaving all other settings at their default values, which implements $\sigma_\epsilon = \sigma_\nu = 1$, then `OK`. The model is just `Ya` and `Za`, with no lags or constant term, then `OK` and `OK` (to skip the next dialog).

In the `Monte Carlo Design` dialog, set M=1000, then `Sample size` as 20:[5]100. Next, click `Monte Carlo Output` and tick `Coefficients` and `Standard errors`, then click `Live Graphics` and tick `Histograms of estimates` and `Recursive Mean of estimates` (the graph below also selected `Standardized data` and edited the outcome), and perhaps set `Plot frequency` to 100, then `OK`. Save the file as ESEBias1.ox, then `OK` to run.

The recursive estimates will appear on screen, and the (edited) graph produced should look like Figure 4.9.

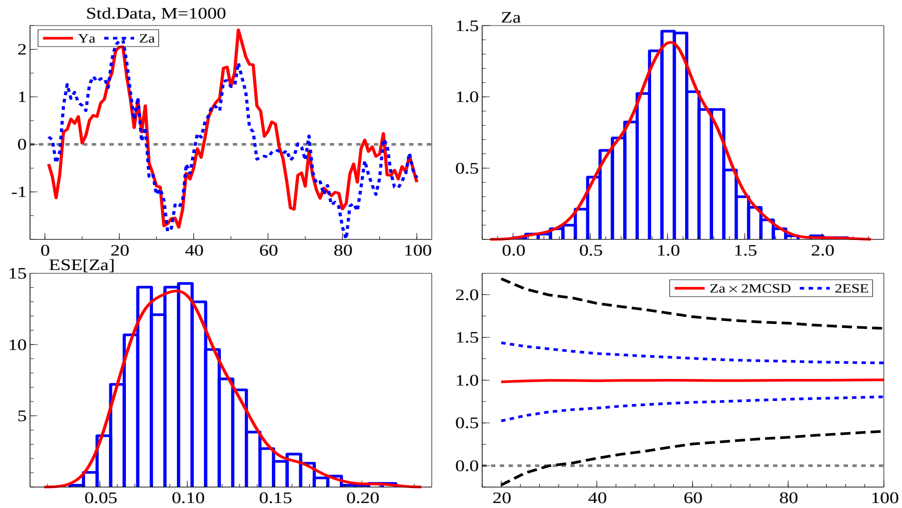


Figure 4.9
Monte Carlo recursive output for mis-specified autocorrelated errors

The two variables are clearly highly correlated, and the distributions of $\widehat{\beta}$ (shown as Za), and its ESE seem fine, with the former centered correctly on unity. It is the final panel that reveals the problem: the MCSD is much larger than the ESE at all sample sizes, about three fold as the output in Table 4.1 confirms, with the MCSD at $T = 100$ being 0.30 as against the mean ESE of just under 0.1. Task 28 will explain why this result occurs.

moments of estimates	mean	MCSD
Za	1.0040	0.30087
ESE[Za]	0.098909	0.029186

Table 4.1
Monte Carlo moments of estimates

A new concept is the MCSD of an ESE. This measures the variation across the Monte Carlo of the ESEs obtained in each of the M replications, so when divided by $\sqrt{M}$, represents the uncertainty with which the mean ESE is estimated in the simulation. As $M = 1000$ here, $\sqrt{M} \approx 32$ so the uncertainty in estimating the mean ESE of roughly 0.1 is just under 0.001.

Repeat the experiment first with $\rho = 0$, $\lambda = 0.9$, then with $\rho = 0.9$, $\lambda = 0$ to check that in both cases the MCSD is close to the ESE: both the error and the regressor need to be autocorrelated.

4.17 Chapter 4 exercises

1. Estimate the regression of real-wage inflation, $\Delta(w-p)_t$, on the unemployment rate, $U_{r,t}$. Discuss the coefficient estimate and evaluation statistics.

2. Now estimate the regression of $(w - p - g + l)_t$ on $U_{r,t}$ and $(w - p - g + l)_{t-1}$. Discuss the coefficient estimates and mis-specification statistics.
3. Can you compare the outcome with that from the model of $\Delta(w - p)_t$ in *i*)?
4. Estimate the regression in *ii*) for sub-samples before 1914; between 1919 and 1939; and after 1945. How would you test if the observed variation over time was due to parameter change or sampling variation?
5. Discuss the difference between a DGP like $\mathsf{D}_{\mathsf{y}}(y_1 \ldots y_T | z_1 \ldots z_T, \theta)$ and an econometric model like $\mathsf{f}_{\mathsf{y}}(y_1 \ldots y_T | z_1 \ldots z_T, \beta)$.
6. Are all models of $\mathsf{D}_{\mathsf{y}}(\cdot)$ equally useful? Can mis-specification tests help distinguish good models from bad?

Chapter 5
Modelling UK unemployment

Chapter 5 guide posts

1. Economic-theory models rarely allow for the sudden shifts seen in data, but may still help inform a statistical model specification: Section 5.1.
2. Theory can guide the formulation of a general unrestricted model (GUM) provided the GUM reflects previous evidence and institutional knowledge and incorporates dynamics, non-linearities, breaks and trends: Section 5.2.
3. A linear dynamic equation can be transformed to an equilibrium-correction model (denoted EqCM), which eliminates stochastic trends by cointegration: Section 5.3.
4. Multiple location shifts can be tackled by impulse-indicator saturation (denoted IIS), with little impact when there are in fact no shifts or outliers: Section 5.4.
5. The resulting GUMs can be too large for humans to handle, and may have more variables than the sample size, but automatic model selection software can handle such GUMs, and still find a parsimonious congruent representation that can also explain all the intermediate results (called encompassing): Section 5.5.
6. An EqCM model of unemployment describes the evidence reasonably well, and remains constant over 2005–2011 despite the turbulence of the 'Great Recession': Section 5.6.
7. That model is not a 'causal' explanation, as a key regressor is contemporaneous, with components that are undoubtedly affected by unemployment, but still has some economic policy implications: Section 5.7.

5.1 Postulating better models

Poverty was perhaps the greatest problem that would face you living in 1860, with its attendant problems of malnourishment and disease, but it would be exacerbated by unemployment, as there was essentially no other source of income for the majority of the population than their earnings (or the Workhouse). Garraty (1979) estimates that the unemployed have been with us even since the Middle Ages, then in the form of sturdy beggars. As Figure 3.7 showed, unemployment rates fluctuated considerably before the First World War, and could reach 10% of the labour force, but that was dwarfed by the persistently high levels in the interwar period, a phenomenon that regrettably returned after the benign post-war reconstruction (and Keynesian) era with the Oil Crises of the 1970s and Mrs Thatcher's economic policies. It is obviously important to understand the behaviour of unemployment much better than the simple models in Chapter 4.

Progress in understanding any economic relationship requires better models, which here could be based on either or both of:
[A] better economic theories of aggregate unemployment;
[B] more general empirical models that tackle stochastic trends, breaks, dynamics, non-linearities and interdependence.
We discussed [A] in section 3.7. $U_{r,t}$ is the outcome from the supply of and demand for labour, aggregated across all prospective workers (L_t^s), facing demands for many different types of workers from all companies and the public sector (L_t^d). In turn those labour demands depend on demands at profitable prices for the goods and services produced. The DGP implicit in such a framework is far too complicated to model in detail here, so instead, we will turn to [B] and relate $U_{r,t}$ directly to aggregate demand and supply. One way of doing so is to use the 'gap' between potential output, g_t^*, and actual, g_t as an approximation to $L_t^s - L_t^d$. Such an approach is quite widely used, but g^* depends on the potential flows of capital services worked on by quality-adjusted hours of available labour inputs, both of which are essentially unknown.

5.1.1 Measuring the gap

Sometimes, $g_t - g_t^*$ is calculated using deviations of g_t from its 'trend' growth, but we have already seen that the growth rate is not constant, which can lead to very mistaken estimates of periods of excess demand and supply.

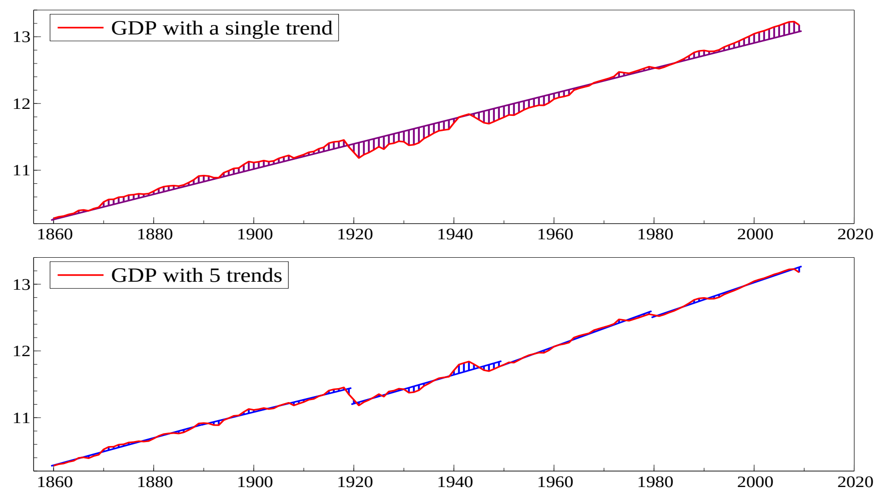


Figure 5.1
Log GDP with one and 5 trends.

The top panel shows one trend line, revealing that g_t was 'above trend' from 1860–1920 and 1980–2004. Such an outcome is sometimes interpreted as excess demand, with being below suggesting excess supply, although nothing in the historical record would support such interpretations. Rather the converse, as the latter part of the 19th Century was once known as the Great Depression till that nomenclature was switched to the 1930s. Certainly the deviations are below trend during the 1920s and 1930s but remain so during the post-WWII reconstruction boom when the unemployment rate was often less than 1%. Thus, deviations from trend are an unreliable guide to excess demand or supply when the trend is not constant.

The lower panel has five separate trend lines, and the much smaller residuals show that the trend changed several times—and consequently this measure of the 'gap' assigns different periods to excess demand and supply. Notice that choosing different sample periods will therefore lead to different assignments of excess demand or supply. As the rates of technical progress over time, and changes therein, are unknown, and the choice of five breaks in trend is arbitrary, we will pursue an alternative approach. However, Chapter 8 returns to the issue of modelling excess demand for goods and services.

5.2 An alternative empirical model of unemployment

Instead of the 'gap', we will use a measure related to it, albeit rather more descriptive than causal of unemployment: see Hendry (2001). This measure assumes that unemployment will fall when hiring labour is profitable, and will increase if it is not profitable. While there is also little accurate data on aggregate profitability, there is a possible 'proxy'—namely an observable variable that is usually reasonably closely related to profits. On the demand side, changes in aggregate revenues are reflected by changes in GDP, namely Δg_t. On the cost side, we saw above the close connection between $(w-p)_t$ and $(g-l)_t$, so average real labour costs are matched by productivity. That leaves capital costs, and while not well measured, those depend on real borrowing costs, which can be taken as $(R_L - \Delta p)_t$. Combining these steps, we approximate the behaviour of changes in real profits by the differential $d_t = ([R_L - \Delta p - \Delta g])_t$, using the negative to have the same time shape as $U_{r,t}$, recorded in Figure 5.2.

The match is less close after WWII, so other factors probably matter as well, which we will consider later when allowing for location shifts.

5.2.1 Modelling unemployment by the profits proxy

Since the paths of the two time series, d_t and $U_{r,t}$, have some common movements, we will model $U_{r,t}$ over $T = 1862 - 2004 by an ADL like$(4.39), using d_t for z_t:

$$\begin{aligned}
\widehat{U}_{r,t} &= \underset{(0.002)}{0.007} + \underset{(0.035)}{0.86}\, U_{r,t-1} + \underset{(0.024)}{0.24}\, d_t - \underset{(0.027)}{0.10}\, d_{t-1} \\
\mathsf{R}^2 &= 0.88 \;\; \widehat{\sigma}_\epsilon = 0.013 \;\; \mathsf{F}_{\mathsf{ar}}(2,137) = 3.08^{*} \;\; \mathsf{F}_{\mathsf{arch}}(1,141) = 2.44 \\
\chi^2_{\mathsf{nd}}(2) &= 7.16^{*} \;\; \mathsf{F}_{\mathsf{het}}(6,136) = 4.97^{**} \;\; \mathsf{F}_{\mathsf{reset}}(2,137) = 5.53^{**}
\end{aligned} \tag{5.1}$$

As a rise in d_t corresponds to a fall in profits, unemployment rises, so the expected sign of its impact in (5.1) is positive. The fit is better than either previous model,

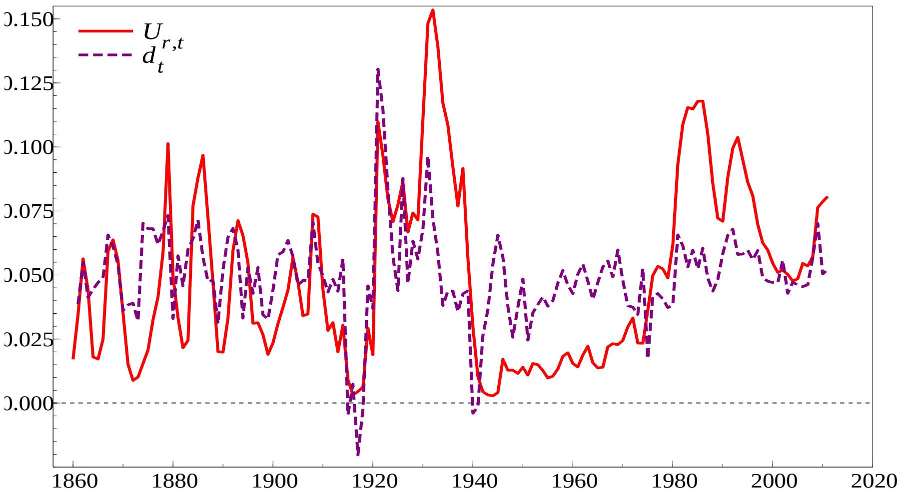


Figure 5.2
Profits proxy and unemployment.

and the impacts of both d_t and its lag are apparently statistically significant on the basis of dividing each estimated coefficient by its standard error to get a statistic that has a t_{139} distribution under the null that each coefficient is zero, using the criterion of lying outside ±2.6 to reject the null hypothesis at the 1% level. However, several of the mis-specification tests discussed in Task 15 Section 4.15 reject their null hypotheses at 1%, so t-tests are possibly unreliable, and a better model should be feasible.

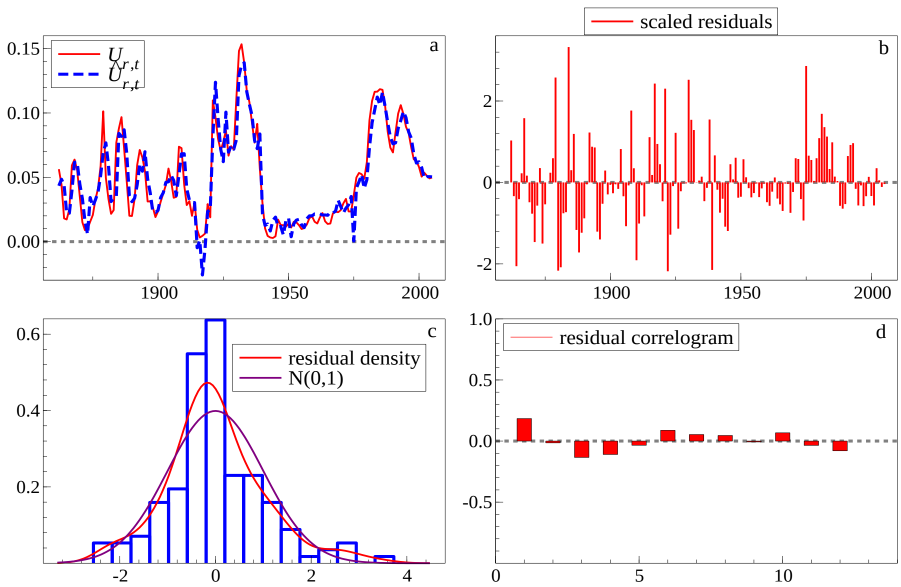


Figure 5.3
Graphical statistics for the dynamic unemployment model (5.1).

Figure 5.3 records the actual $U_{r,t}$ and fitted values $\widehat{U}_{r,t}$ in panel a, the standardized residuals $\widehat{\epsilon}_t/\widehat{\sigma}_\epsilon = (U_{r,t}-\widehat{U}_{r,t})/\widehat{\sigma}_\epsilon$ in panel b, and their density and correlogram in panels c and d respectively. The residuals are close to an innovation, and seem to be reasonably normally distributed, but are somewhat heteroskedastic, and we still need to consider the constancy of the parameters and the validity of conditioning on d_t, issues we must leave for the moment. The model is also not 'data admissible' as it generates negative fitted values of $U_{r,t}$, although that could be 'fixed' by modelling either $\log(U_{r,t})$ or the logistic $\log[U_{r,t}/(1-U_{r,t})]$.

5.3 Interpreting the dynamic unemployment model

According to (5.1), a rise in d_t first raises unemployment, but then appears to lower it in the next period when $-0.10d_{t-1}$ 'kicks in', which seems odd at first sight. However, we can apply the analysis in section 4.10.1 to (5.1) to understand that equation better. Interpreting (5.1) as in (4.39), then $\widehat{\beta}_0 = 0.007$, $\widehat{\beta}_1 = 0.24$, $\widehat{\beta}_2 = 0.86$, and $\widehat{\beta}_3 = -0.10$, so that $\widehat{\kappa}_0 = 0.007/0.14 = 0.05$ and $\widehat{\kappa}_1 = (0.24-0.10)/0.14 = 1.0$ from (4.43).[1] Written as in (4.44), then (5.1) becomes:

$$\Delta\widehat{U}_{r,t} = 0.24\Delta d_t - 0.14\left(U_{r,t-1} - 0.05 - 1.0d_{t-1}\right) \tag{5.2}$$

so the long-run equilibrium using rounded coefficients is:

$$U_r = 0.05 + d \tag{5.3}$$

or 5% unemployment when $d = 0$, which is its mean.

Unemployment rises or falls by approximately 1% for every 1% decrease or increase in $d = (R_L - \Delta p - \Delta g)$. The immediate effect of a change in d is an impact of ±0.24%, so unemployment only moves part of the way to the eventual impact of 1% and that creates a disequilibrium. Then, 14% of that deviation from equilibrium is removed each period. Thus, the apparently hard to interpret positive effect from d_{t-1} transpires to be the natural consequence of an equilibrium correction.

That interpretation requires that $(1-\widehat{\beta}_2)$ and $(\widehat{\beta}_1+\widehat{\beta}_3)/(1-\widehat{\beta}_2)$ are significantly different from zero, allowing for the variables to be non-stationary, which entails larger critical values than conventional (see e.g., Banerjee, Dolado, Galbraith, and Hendry, 1993). In fact, the *PcGive* t-test for a unit root in (5.1) indeed rejects, so the equilibrium in (5.3) is $\mathsf{I}(0)$: see Task 18 Section 5.10.

5.3.1 Allowing for longer lags

Although the dynamic model in (5.2) is now interpretable, it has an important restriction—we only allowed for one lag, so excluded any effects from lagged changes like $\Delta U_{r,t-1}$ and Δd_{t-1} (or longer). Those are easily added to equations like (5.2), and doing so delivers:

$$\begin{aligned}
\Delta\widehat{U}_{r,t} &= \underset{(0.07)}{0.16}\,\Delta U_{r,t-1} + \underset{(0.02)}{0.24}\,\Delta d_t - \underset{(0.03)}{0.12}\,\left(U_{r,t-1} - 0.05 - d_{t-1}\right) \\
(\mathsf{R}^*)^2 &= 0.48\quad \widehat{\sigma}_\epsilon = 0.012\quad \mathsf{F}_{\mathsf{ar}}(2,137) = 0.32\quad \mathsf{F}_{\mathsf{arch}}(1,140) = 3.25 \\
\chi^2_{\mathsf{nd}}(2) &= 16.2^{**}\quad \mathsf{F}_{\mathsf{het}}(6,135) = 5.77^{**}\quad \mathsf{F}_{\mathsf{reset}}(2,137) = 3.50^{**}
\end{aligned} \tag{5.4}$$

[1] These use rounded estimates: when the actual values are used, $\widehat{\kappa}_1 = 1.05$.

As can be seen, $\Delta U_{r,t-1}$ is significant when added, but Δd_{t-1} was not, so was eliminated. Since (5.4) does not have a free intercept, $(\mathsf{R}^*)^2$ is calculated when a constant term is added. The resulting much smaller value than in, say (5.1), is because the dependent variable is $\Delta U_{r,t}$, rather than $U_{r,t}$: $\widehat{\sigma}_\epsilon$ is smaller, so the model is definitely an improvement, and in turn R^2 is revealed to be an unreliable measure of 'goodness of fit'.

The effect of $\Delta U_{r,t-1}$ is to increase the inertia in unemployment, adding to rises as unemployment increases. The graphical statistics for (5.4) are shown in Figure 5.4. Although the match of d_t and $U_{r,t}$ seemed best in the 19th Century in Figure 5.2, the fitted values track the outcomes least well for the changes in unemployment over that period. However, the unemployment data then were based almost entirely on Union members. The resulting significant rejections of Normality, homoskedasticity and linearity are possibly due to that change in measurement, but will be addressed shortly. Notice that congruence in even a greatly improved model can still be rejected, re-emphasizing the drawbacks of a strategy of augmenting overly simple initial models.

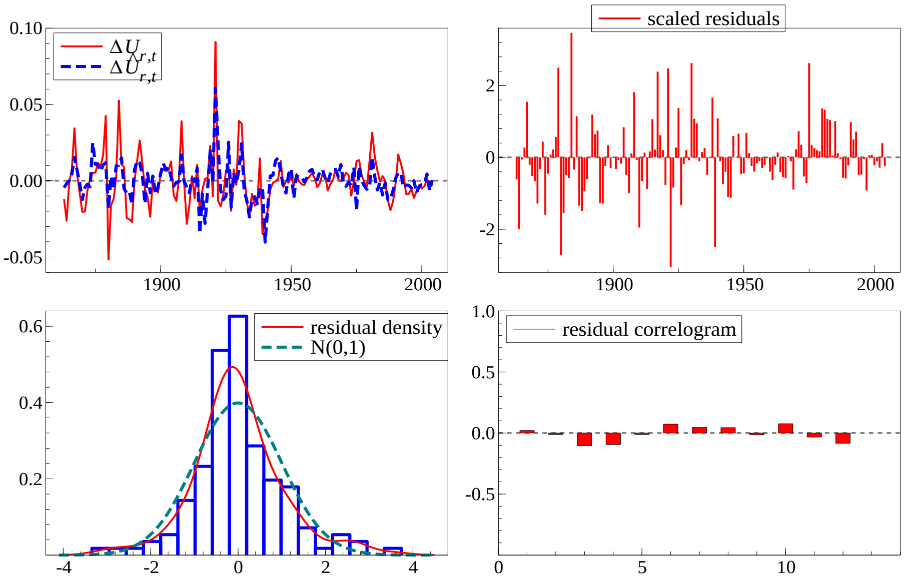


Figure 5.4
Graphical statistics for the equilibrium correction unemployment model (5.4).

5.4 Outliers and location shifts

Earlier we noted four mean shifts in $U_{r,t}$, but indicators for these were not included when estimating (5.4). Moreover, a couple of 'outliers', or large scaled residuals, defined as outside ±2, are visible in panels *b* and *c* of Figure 5.4. Both features need to be included, albeit the former will not be needed when there is co-breaking between $U_{r,t}$ and d_t, but the latter are a slight danger signal as follows. Let:

$$y_t = \mu_0 + (\mu_1 - \mu_0)\, 1_{\{t \geq T_1\}} \tag{5.5}$$

so there is a location shift at time T_1. Then, differencing:

$$\Delta y_t = (\mu_1 - \mu_0)\,\Delta 1_{\{t \geq T_1\}} \tag{5.6}$$

As $\Delta 1_{\{t \geq T_1\}}$ must be zero when the indicator is zero, which happens for all $t < T_1$, and must also be zero when the indicator is unity, as occurs for all $t > T_1$, then it is only non-zero at the switch point, $t = T_1$. Thus, the change in the mean-shift indicator is an 'impulse' indicator, $1_{\{t=T_1\}}$:

$$\Delta y_t = (\mu_1 - \mu_0)\,1_{\{t=T_1\}} \tag{5.7}$$

Consequently, finding impulse indicators in an equilibrium-correction model can entail that a location shift has occurred but has been differenced.

However, care is required in how one checks for outliers and shifts. Figure 2.12 illustrated a location shift in the artificial data series generated by (2.12), so $y_t = \mu_0 + (\mu_1 - \mu_0)\,1_{\{t \geq T_1\}}$ where the shift was $\mu_1 - \mu_0 = 5$ error standard deviations, which occurred at $T_1 = 0.23T = 23$ from $\mu_0 = 15$. This graph is reproduced as Figure 5.5.

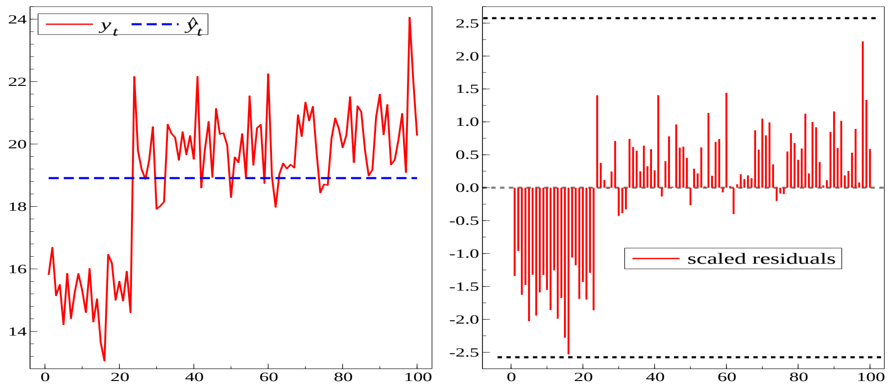


Figure 5.5
Location shift with no outliers at 1%.

The right-hand panel now records the scaled residuals that result if the fitted model is just a constant as shown in the left-hand panel: the estimated mean is the weighted average of $(0.22 * 15 + 0.78 * 20) = 18.9$, so there is a large residual standard deviation. Thus, despite the large shift of 5 standard deviations, there are no outliers at 1%, as all residuals lie in the range $\pm 2.575\widehat{\sigma}$. The absence of large residuals need not entail the absence of shifts. This example has also been included to highlight the dangers of not looking at the model's graphics, which clearly reveal the shift and the mis-representation of the evidence by a single mean.

5.4.1 Allowing for outliers and location shifts in $U_{r,t}$

A first test of (5.4) is to add the four location-shift indicators for 1860–1913, 1914–1938, 1939–1979, and 1980 onwards, noting that there is no free intercept to cause perfect collinearity problems (because a linear combination of the indicators equals the constant, there would be two intercepts if (5.4) already included one). That addition yields t-values of −1.7, 1.94, −0.90, and 1.1, none of which rejects the null that their effect is zero, nor does a test of their joint significance reject. Thus, (5.4) seems to have 'captured' those four location shifts, evidence of co-breaking between $U_{r,t}$ and d_t. However, given the lesson from (5.7), we next check for outliers.

Problematic indicators would correspond to the years when the jumps occurred, namely 1914, 1939, and 1980. One could just add those impulse indicators, but there may be other shifts relative to this model that derive from the presence of d_t, which might mask the significance of any or all of those three. A general approach would be to allow for possible impulse indicators at every data point to check for both location shifts and outliers. At first sight, that may seem impossible, as doing so would involve T indicators for T observations and surely lead to a perfect fit. Surprisingly, however, it can be done, and is called impulse-indicator saturation, abbreviated to its acronym IIS. While it is beyond the level of this introduction to offer a detailed theoretical explanation, we now discuss how it is statistically feasible to allow for T indicators and additional variables for T observations. The distributions of the indicators and the resulting parameter estimates have been derived for models of the type we are estimating here (see Hendry, Johansen, and Santos, 2008, and Johansen and Nielsen, 2009); and IIS has been programmed as an option in *PcGive* (called *Autometrics*: see Doornik, 2009).

5.5 Impulse-indicator saturation

The logic behind adding a complete set of impulse indicators, $\{1_{\{t\}}\}$ for $t = 1, \ldots, T$, to a model selection procedure can be understood by considering two stages, each of which uses feasible subsets of $T/2$, known as the split-half approach: see Hendry and Nielsen (2010). An impulse indicator at observation τ ensures a zero residual $\widehat{\epsilon}_\tau = 0$. Including the first half of T indicators therefore ensures a perfect fit over that sub-sample, and is said to 'dummy out' those observations, i.e., is equivalent to dropping them from the estimation sample. Consequently, only the second-half data are used to estimate any other parameters. Under the null that there are no outliers or shifts, such estimates will be unbiased, although inefficient, which also holds for $\widehat{\sigma}$. Now checking the significance of the indicators from the first half can reveal which observations are discrepant. On replacing the first half by the second, that analysis holds again; so we have successfully checked for shifts and outliers at all observations.

A basic probability result is that when testing N independent null hypotheses at a significance level α, then on average $N\alpha$ will reject by chance. For example, if $N = 100$ and $\alpha = 0.01$, one hypothesis is likely to be falsely rejected. Somewhat surprisingly, such a result continues to hold for testing T hypotheses, as with impulse indicators which are mutually orthogonal. Since there are T impulse indicators used in IIS, it seems natural to set $\alpha = 1/T$, so on average one irrelevant indicator will be retained by chance from a distribution with no outliers or location shifts.

Figures 5.6 and 5.7 illustrate how impulse-indicator saturation works, based on the split-half approach, first seeing how it would perform under the null when there is no shift and then when the shift in Figure 5.5 is present.

The outcome under the null when there is no location shift is shown in Figure 5.6. The left-most column records which impulse indicators were included; the middle column shows which impulse indicators were significant; and the right hand column records the resulting fitted and actual values when those indicators

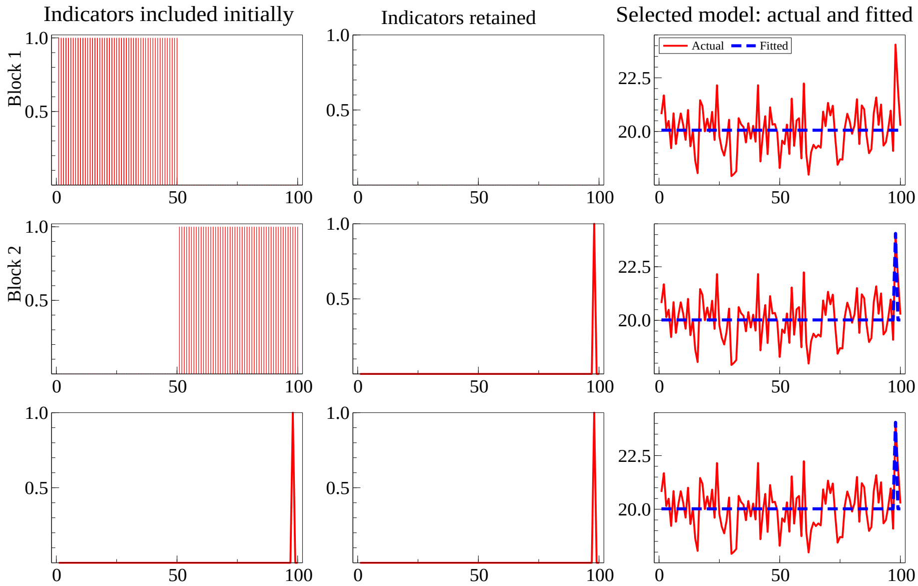


Figure 5.6
Split-half IIS under the null.

were retained even though none is needed. The top row is when the first set of $T/2$ impulse indicators are included; the middle row is the second set after the first set are dropped; and the bottom row is when the significant impulse indicators from those two stages are combined, and re-selected for significance. As can be seen, when there is no break, one impulse indicator is retained when undertaking individual t-tests at a 1% significance level, which is what would be expected to happen on average under the null when $T = 100$ indicators are tested (see Castle, Doornik, and Hendry, 2012, for a discussion of applying IIS to non-normal distributions).

Figure 5.7 records the outcome when there is a location shift. The top row shows that 23 impulse indicators are retained initially matching the shift, as the mean is shifted by 5 after $T = 23$. The second row shows that none of the additional indicators are retained: the large late onset outlier is missed as the error standard deviation is inflated by not including the indicators from the first period. The third row demonstrates that combining the impulse indicators selected at the first two stages and re-selecting need not alter the outcome (but can do), as here only those from the first stage matter. As these all have similar magnitudes and the same sign they could be combined into a single location-shift indicator (see Hendry and Santos, 2005). In practice, IIS is an option in *Autometrics*, which undertakes many searches cumulating 'knowledge' about the process under analysis before terminating when no further outliers can be found and the model is the most parsimonious congruent representation. Here, IIS locates the later outlier as well as retaining the first 23 indicators.

IIS has surprising implications. Despite testing for any number of location shifts and outliers anywhere in the sample by adding T impulse indicators in large blocks, the cost under the null is eliminating one observation at $\alpha = 1/T$ from the single impulse indicator retained on average by chance. That is a tiny cost against the potential benefits of removing breaks that could distort inference. Further, IIS

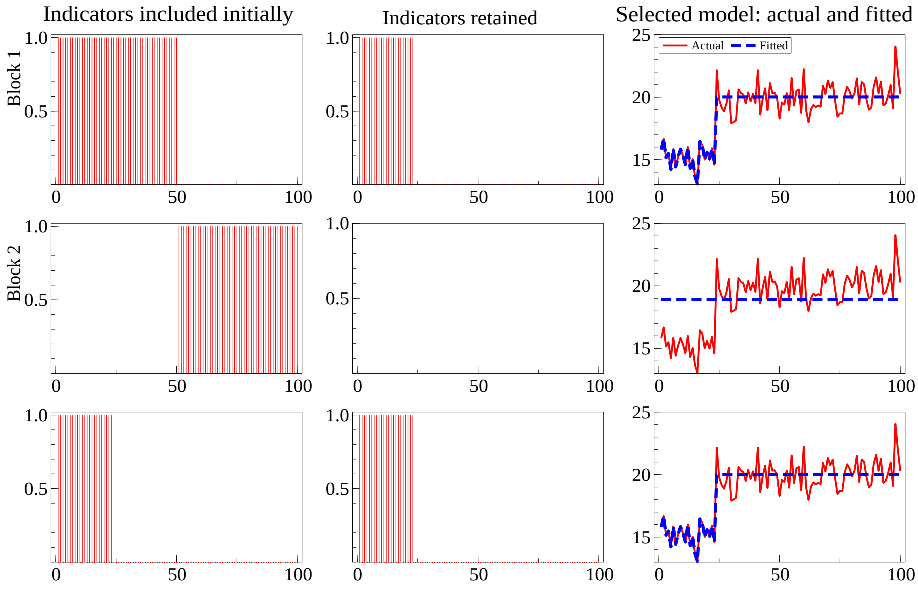


Figure 5.7
Split-half IIS for a location shift.

demonstrates that having more variables than observations need not be an infeasible setting when handled appropriately.[2]

5.5.1 IIS for the unemployment model

Rather than just the two half-samples, *Autometrics* uses repeated blocks of multiple sample splits, such as $T/3, T/4, T/5$, usually selecting relevant indicators at a tight significance level such as 1% or 0.5% depending on the sample size. We can apply *Autometrics* to check for outliers in (5.4) at every data point using IIS at the very tight significance level of 0.1%, so there is just one chance in 1000 of finding an impulse where there was no genuine outlier: all the economic variables were retained while doing so. This yielded the following significant impulse indicators, with their magnitudes shown in parentheses: 1879 (3.0%), 1880 (−5.0%), 1884 (4.5%), 1908 (2.8%), 1921 (5.3%), 1922 (−5.1%), 1930 (3.5%), and 1939 (−3.6%). Consequently, several early outcomes, the 1921–1922 crash, and the start of the Great Depression in 1930 are not captured by the model in (5.4). However, only 1939 matches the date for a danger signal, as it is close to the timing and magnitude of the third location shift ($\mu_3 - \mu_0$) of about −3%. Thus, the model may not be explaining the low unemployment over the post-war reconstruction era, although that is not a feature that can be discerned in Figure 5.8.

[2] A variant of this technique, called step-indicator saturation–which uses step, rather than impulse, indicators described in Castle, Doornik, Hendry, and Pretis (2015)–was used to estimate the parameters of (1.6) at a 0.1% significance level: see Section 6.2.1.

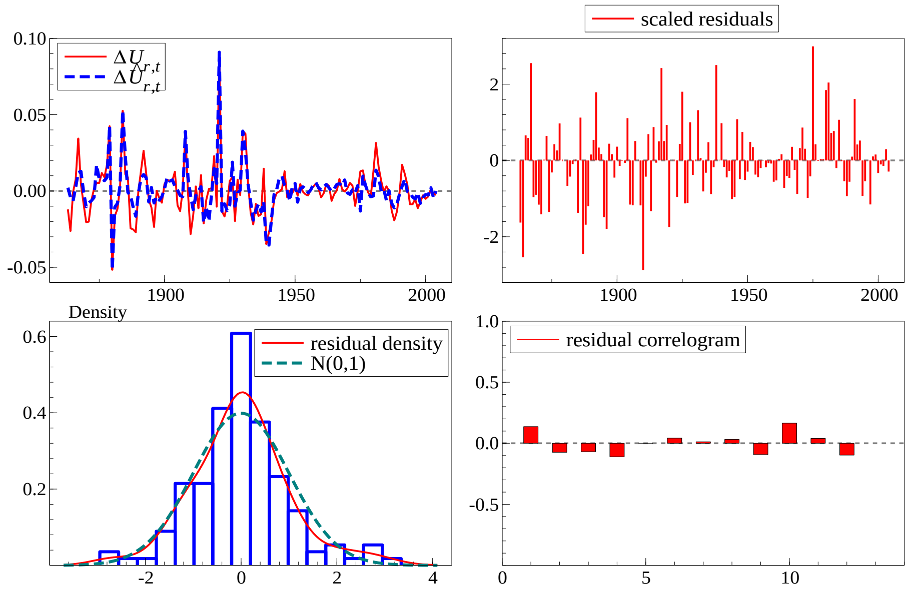


Figure 5.8
Graphics of the EqCM unemployment model with IIS.

5.6 'Forecasts' from the EqCM unemployment model

Models like (5.4) estimated above were in fact developed a few years ago. Since then, there have been seven new annual outcomes on all the variables. The period includes the sharp rise in unemployment during 2008, which nevertheless was much smaller than expected given a fall of more than 6% in real GDP. A second check on (5.4) is to 'forecast' these recent outcomes. Such 'forecasts' are just the fitted values of (5.4) over 2005–2011, not genuine statements about the future, which would require forecasts of d. We use the estimates after IIS which were:

$$\begin{aligned}\Delta\widehat{U}_{r,T+h|T+h-1} &= 0.16\Delta d_{T+h} + 0.35\Delta U_{r,T+h-1} \\ &\quad -0.08\left(U_{r,T+h-1} - 0.05 - d_{T+h-1}\right) \qquad (5.8)\end{aligned}$$

where $T = 2004$ and $h = 1, \ldots, 7$, so 'forecasting' 1-step at a time from known current and past outcomes.

Figure 5.9 records the findings of that check for both the levels and changes in the unemployment rate, and shows that the estimated model tracks the new data accurately. The top panel reports the whole history with the fitted values and forecasts, which are shown with error bands estimating the uncertainty that should include the outcomes 95% of the time when the new data are consistent with the estimated model. The lower panel for the changes show 95% vertical bars. Both sets of forecast intervals easily include all the outcomes. In effect, this is a test of the constancy of the model over a turbulent period, when the increase in unemployment was far less than might have been expected for the very large fall in GDP.

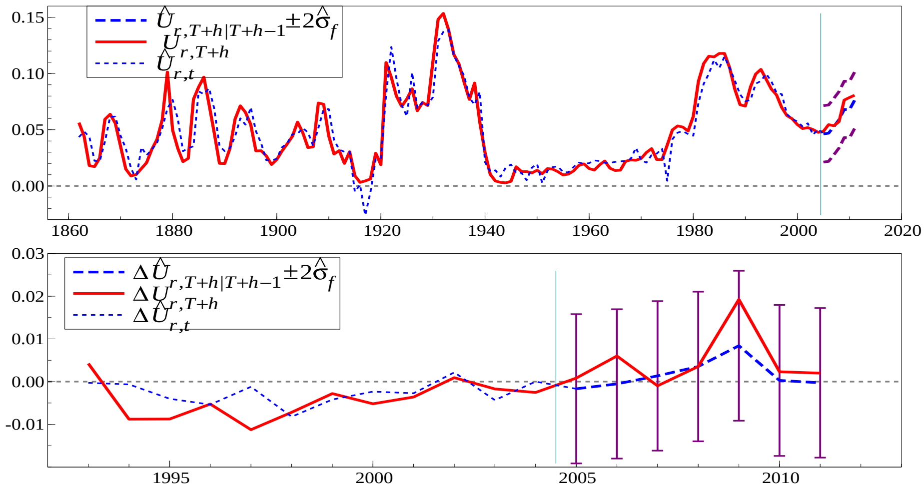


Figure 5.9
‘Forecasts’ from the EqCM unemployment model (5.8).

5.7 Some policy implications of the unemployment model

Equation (5.4) entails a number of policy implications as follows.
(i) When the real long-term interest rate, $R_L - \Delta p$, equals the real growth rate, Δg, so $d = 0$, equilibrium unemployment is about 5%, close to the average historical unemployment rate. The model does not explain why 5% occurs, merely that movements from 5% are associated with non-zero values of d.
(ii) To permanently lower U_r below 5% would require lowering d, either by reducing real long-run interest rates, or raising the long-run growth rate of GDP. Both are policies the government favours, but the latter is difficult to maintain while imposing austerity, especially when that reduces education of the next generation.
(iii) Unemployment can be well below its equilibrium for long periods when $d < 0$ also holds, as occurred over 1939–1968. A key policy issue is the possible consequence of that for inflation, namely did the persistently low unemployment rate during the post-war reconstruction era lead to the high inflation of the 1970s, or did that have a different cause? We turn to modelling wage inflation next, after summarizing some of the lessons of this chapter.

5.8 Chapter 5 key points

(A) Economic-theory models rarely allow for sudden shifts, so often fail to adequately characterize macroeconomic data, but may still help inform a statistical model specification.
(B) To represent the DGP at all accurately, one must start an empirical analysis from a general unrestricted model (GUM) that includes all the substantively relevant variables suggested by economic theories, previous evidence and institutional knowledge, as well as their dynamics, possible non-linearities, breaks and trends.

(C) Dynamic equations can be transformed to equilibrium-correction models (denoted EqCMs), which eliminate stochastic trends in cointegrated relations, and make interpretation easier.
(D) Multiple location shifts can be tackled by indicator saturation, which has little impact on an empirical modelling exercise when there are no shifts or outliers.
(E) The general unrestricted models that result from (B) and (D) will not only be too large for humans to handle, they will usually contain more variables plus indicators than the sample size.
(F) Nevertheless, as illustrated for some simple IIS examples, automatic computing software can handle such GUMs, and reduce the initial proliferation to a parsimonious, congruent and encompassing selection.
(G) The unemployment EqCM model based on $d_t = (R_L - \Delta p - \Delta g)_t$ describes the evidence relatively well, and is constant on the new data over 2005–2011 despite the turbulence of the 'Great Recession'.
(H) That model is not a 'causal' explanation, as d_t is contemporaneous and its components are almost surely affected in turn by the level of unemployment.
The next chapter will apply the methodology we have developed so far to model real wages, then Chapter 7 will consider modelling UK money demand, and finally Chapter 8 will investigate price inflation.

5.9 Task 17: Estimating a dynamic equation

Create and estimate the dynamic model of *Ur* in (5.1): `Formulate`, and `Clear` any previous model selection if necessary, set `Lags` to one, and double click on *Ur*, then *d*, then `OK`, `OK`, and if $T = 1862 - 2004$, then `OK` to estimate. Its 4-quadrant graphs follow as in Task 14 Section 4.14. Discuss its mis-specification test outcomes, and compare these to earlier estimated equations for U_r.

5.10 Task 18: Estimating an equilibrium-correction equation

PcGive can solve for the implicit equilibrium of (5.1), and test that the relation is a cointegrating one as follows. In the `Test Menu` tick the `Dynamic Analysis` box, then `OK` to bring up its dialog. Tick the boxes for `Static long-run solution` and `Lag structure analysis`, then `OK`. The former solves for the equilibrium as in section 5.3, and the second provides the *PcGive* t-test for a unit root (and hence for cointegration when the variables are I(1)). The solved equilibrium should be as in (5.3) when rounded, and the `Unit-root t-test` is $t_{ur} = -4.03^{**}$ so rejects at 1% (see Ericsson and MacKinnon, 2002). Although both $U_{r,t}$ and d_t are non-stationary, neither is integrated, but the procedure can still be applied and confirms the absence of a unit root.

Round the expression 'ECM = Ur - 0.0492471 + 1.04499*d;' and rename to 'ECMUrd = Ur - 0.05 + d;' highlight and Ctrl+**A** to calculate and add to the end of the database. To `Formulate` (5.4), `Clear` then with one lag set, double click on *DUr*, then *Dd* and finally *ECMUrd*. Click inside the `Selection` window and delete the current-dated *ECMUrd* but keep its lag (the contemporaneous value would lead to a perfect fit as it is just a linear combination of $U_{r,t}$ and d_t). `OK`, `OK`, then set the sample to 1863 – 2004. The estimates should show that the Constant and Dd_1 are insignificant, so return to formulate, remove them and re-estimate to replicate (5.4).

You have just done manual model selection: let's compare that with automatic selection. Return to `Formulate`, and at the foot of the `Selection` window, click on the down arrow at `Recall a previous model` to reinstate the general equation with the Constant and Dd_1 included. `OK`, but this time tick the box for `Automatic model selection`, set `Target size` to `Medium: 0.025`, accepting the remaining default options, `OK` then set the sample to 1863 – 2004 and `OK`. The outcome might surprise: the selected model retains Dd_1 even though it is insignificant. The explanation is that by doing so, *Autometrics* substantively reduces the magnitude of the heteroskedasticity tests, as you can see on comparing with your simplified re-estimation of (5.4). The selection procedure is predicated on starting from a congruent model, which is not the case here, so the terminal choice can reflect that. We now need to address the non-congruence.

5.11 Task 19: Impulse-indicator saturation

To tackle the issue of changes in measurement and possible outliers and location shifts, we need to use impulse-indicator saturation as discussed in section 5.5. Fortunately, this is easy. `Formulate` and reinstate the general model for Task 18 with 5 regressors. Highlight these, right click with the mouse and select `U: Unrestricted (fixed)`, so that all will be retained while searching by IIS for shifts etc. `OK`, and again tick the box for `Automatic model selection`, setting `Target size` to `Tiny:0.001` as we only want important shifts removed; and double click on `None` opposite `Outlier and break detection`, setting that to `Impulse-indicator saturation (IIS)`, then `OK` and `OK`, after which calculations may take some time. This should replicate the results in section 5.5.1. Eliminating both the insignificant Constant and Dd_1 now *improves* the diagnostic statistics, with only Normality in doubt: the very stringent significance level for IIS could omit outliers that appeared significant at 5%.

So far, we have only considered the default mis-specification tests (called diagnostics when they are used to guide the selection of congruent models). On the `Test Menu`, those can be calculated by ticking the box for `Test Summary`. However, other tests are available, including the test for non-linearity in Castle and Hendry (2010). On the `Test Menu` tick the box for `Test ...` and then tick the box for `Index test for non-linearity` on the `Test dialog`. The output provides two versions, which are usually similar, and here the second (Core index test) delivers 'F(9,122) = 1.20', which does not reject.

5.12 Task 20: 'Forecasting'

The final Task in this chapter is 'forecasting' the seven withheld data points, 2005–2011, which period includes the Financial Crisis and the Great Recession. However, as d_t is a contemporaneous regressor, so could not be known for a genuine *ex ante* forecast, the exercise is really one of checking parameter constancy outside the estimation sample. Indeed, as the Financial Crisis was essentially unanticipated a year ahead, a crystal ball would be needed for a genuine forecast.

Express (5.8) in a general notation where Δy_{T+h} is to be forecast successively 1-step ahead over a period $h = 1, \ldots, H$ from a forecast origin at T with one known regressor x_{T+h} using estimated parameters $\{\widehat{\beta}_i,\ i = 1, \ldots, 3\}$, where the equilibrium-correction estimated parameters are $\{\widehat{\kappa}_i,\ i = 1, 2\}$ as:

$$\Delta \widehat{y}_{T+h|T+h-1} = \widehat{\beta}_1 \Delta x_{T+h} + \widehat{\beta}_2 \Delta y_{T+h-1} + \widehat{\beta}_3 \left(y_{T+h-1} - \widehat{\kappa}_0 - \widehat{\kappa}_1 x_{T+h-1} \right) \tag{5.9}$$

At each time $T + h - 1$, the right-hand side is known, so $\Delta \widehat{y}_{T+h|T+h-1}$ can be calculated.

The 'forecast' errors are the deviations $\widehat{u}_{T+h|T+h-1} = \Delta y_{T+h} - \Delta \widehat{y}_{T+h|T+h-1}$ from which statistics such as root mean square forecast errors (RMSFEs) can be calculated, where RMSE= $\frac{1}{H} \sum_{h=1}^{H} \widehat{u}^2_{T+h|T+h-1}$. Forecasts for the level y_{T+h} can be derived from:

$$\widehat{y}_{T+h|T+h-1} = \Delta \widehat{y}_{T+h|T+h-1} + y_{T+h-1}.$$

The RMSFE for y_{T+h} forecast by $\widehat{y}_{T+h|T+h-1}$ is calculated by:

$$\frac{1}{H} \sum_{h=1}^{H} (y_{T+h} - \widehat{y}_{T+h|T+h-1})^2.$$

Select `Model`, `Formulate` and keep the model from Task 19, `OK` and `OK`, then set the sample to 1863 – 2011 and change `Less forecasts` from 0 to 7. The estimated equation should be unaltered, with some new output shown below `1-step (ex post) forecast analysis 2005-2011`. The main statistic of interest here is 'Chow F(7,131) = 0.34077 [0.9338]' which does not reject the null of parameter constancy (see Chow, 1960).

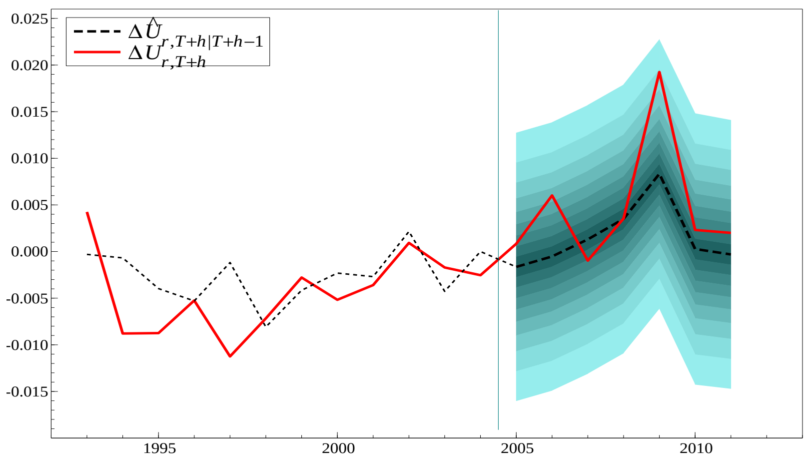


Figure 5.10
Fan chart 'forecasts' for ΔU_r.

A more detailed analysis is provided by selecting `Test Menu`, and ticking the box for `Forecast...` to bring up its dialog. Tick `h-step forecasts`, making sure that h=1 is set. Next, double click on `Error variance only` under `Forecast standard errors` to change that to `With parameter uncertainty`, so that interval forecasts reflect the parameter estimation variances. Those intervals can be represented graphically in a number of ways and for different interval magnitudes: the defaults are error bars and 2 standard errors. Click on `Graph Options` to reset the first to `Use error fans`, and `OK` to see Figure 5.10 (after a little editing, where $\widehat{U}_{r,T+h|T+h-1}$ denotes the forecast value of $U_{r,T+h}$ made at time $T + h - 1$).

All the outcomes lie within the intervals of the fan chart, where lighter shading denotes a lower probability region. Given the large fall in GDP, the relatively small rise in unemployment has been considered surprising, but appears consistent with this historical relation, as the drop in R_L and the failure of inflation to fall combined to offset much of the impact of the large negative value of Δg.

The detailed forecast statistics are also written to the Results file, and show that the RMSFE is 0.0051, which is smaller than $\widehat{\sigma}_\epsilon = 0.0087$, so the model fits better over the later data without any impulse indicators, than it did in-sample with indicators, which is a rare occurrence in macro-econometrics.

5.13 Chapter 5 exercises

1. Estimate the regression of $U_{r,t}$ on $U_{r,t-1}$, without the constant, over a sample ending in 2011, with 7 observations retained for forecasting. Discuss the coefficient estimate and the model evaluation statistics.
2. How well does the forecast RMSFE compare to that of the model in Task 20, Section 5.12? Is this a fair comparison, given that the forecasts from *i*) are genuinely *ex ante*?
3. Can you explain why an unsatisfactory model like *i*) can forecast respectably over the Great Recession?

Chapter 6
Modelling UK wages

Chapter 6 guide posts

1. Empirical analyses need to start from GUMs that include all the theory suggested variables, their dynamics, shifts and non-linearities: Section 6.1.
2. Location shifts are reasonably captured by a generalization of IIS called step-indicator saturation (denoted SIS): Section 6.2.
3. A non-linear response of real wages to erosion from inflation shows an increasing reaction at higher inflation, consistent with the historical occurrence of ‘wage-price spirals’: Section 6.3.
4. Viable empirical models of real-wage growth can be developed, which can be transformed to explain the nominal or real level of wages: Section 6.4.
5. There is little evidence that expectations of future values play an important role in the wage model: Section 6.5.
6. There is also a non-linear response to unemployment, with real wages rising as more workers become unemployed, consistent with unemployment being involuntary: Section 6.6.
7. Even starting with many candidate explanatory variables, their lagged values and non-linear functions thereof, the selected relationship is still perturbed by a few large shifts, mainly coinciding with wars, but selecting location shifts by SIS does not preclude finding non-linearities nor do non-linearities capture all the shifts, so both play important roles: Section 6.7.
8. Despite selecting from a GUM with more candidate variables than observations, the final real-wage model is readily interpretable—with a long-run equilibrium of a constant share of wages in GDP embodied in an EqCM that corrects real-wage ‘losses’ from previous incomplete adjustments to changes in inflation, unemployment and productivity—is constant over the Great Recession, encompasses other empirical models, and passes a stringent test of its invariance to regime changes (called super exogeneity): Section 6.8.

6.1 Wage determination theories

Returning to 1860, and how well you might be doing, Mackenzie (1921) records the median, upper and lower quartiles of adult men’s wages in the UK in 1860, 1880 and 1914. She shows ranges of 14s.6d to 22s.6d in 1860, rising to 25s.2d–39s.4d in 1914 (there were 20 shillings to a £and 12 pence, denoted d, to a shilling): these square with the data in Pember Reeves (1913). Your wage would not buy you very much with bread costing about 1.5d per pound, milk 2d per pint, and (e.g.) bacon 1s per pound. Johnson (1988) suggests that over half of total expenditure was allocated to food (Mackenzie, 1921 has the figure nearer to 66%), 20%–30% to rent, about 8% to fuel and light, and between 3% and 7% to clothing. On either estimate, almost no income was left, and malnutrition was rife. We need to understand why 1860 workers were doing so badly on average, and why matters improved so greatly.

Most analyses of wage inflation postulate the labour market as the main source, through:
(i) excess demand for labour;
(ii) competition over the profit share.
Factor markets seem to determine wages, and the prices of capital goods. But factor demands are derived from final demands, so the latter must be the primary determinants of price inflation (addressed in Chapter 8): inflation usually needs adequate final demand to sustain it.

A model of the overall inflation process needs equations for: prices (or the profit markup of prices over costs), wages, excess demands for goods and services, and unemployment. These usually depend on productivity (output per employee), costs (usually unit labour costs or wages relative to productivity), import prices, commodity prices, taxes and benefits, and exchange rates, as well as special factors (such as the price and availability of energy), sometimes financial variables like interest rates and money supply, all with lagged reactions and perhaps expectations of future prices. However, some models of 'inflation' conflate wages and prices.

Policy analyses of 'inflation' have emphasized the 'natural rate of unemployment' or 'non-accelerating inflation rate of unemployment' (NAIRU)—when voluntary unemployment fell below its 'natural rate', it was claimed inflation would accelerate indefinitely: see Friedman (1977), with UK contributions from Nickell (1990), and Layard, Nickell, and Jackman (1991) *inter alia*. The model is almost the opposite of that in Keynes (1936) and Phillips (1958), who stressed the role of *involuntary* unemployment putting downward pressure on wages. Some models of price-setting firms treat wages as 'cost-push', raising final-goods prices: see for example Dicks-Mireaux and Dow (1959) and Godley and Nordhaus (1972).

Wage-price dynamic interactions (spirals) are also important in the literature: see Sargan (1964), Sargan (1980) for the UK, and (e.g.) de Brouwer and Ericsson (1998) for Australia. Rowlatt (1988) is a good summary of the forces driving UK inflation over 1969–85, when it reached high—almost hyperinflationary—levels. Nevertheless, few theory models allow for major unanticipated shifts.

6.2 Wage, price and unemployment location shifts

Chapter 1 described the data on the two nominal variables wages, W, and prices, P. As an economy has only one nominal level, we transform w and p to create one real (or constant price) variable, $w - p$, and one nominal level, p, from which both nominal variables can be recovered if required. Real wages are modelled in this chapter and nominal prices, or rather inflation Δp, in Chapter 8.

We first extend Table 2.2 to Table 6.1 by adding the comparable statistics for shifts in the UK unemployment rate and productivity growth, $\Delta(g-l)$, shown in Figures 3.7 and 1.12. As discussed in Section 3.9, the four shifts in U_r do not match the eight shifts in wage and price inflation, helping explain the unstable Phillips curves in Figure 3.9. More of a surprise is that many of the shifts in $\Delta(g-l)$ do not match those in $\Delta(w-p)$. However, the location shifts in wage and price inflation are closely similar, confirmed by the considerable co-breaking between wage inflation and price inflation shown by real-wage changes having only one large spike in 1940, and a different mean pre and post WWII as seen in Figure 2.3.

Sub-sample	Δw	Δp	U_r	$\Delta(g-l)$	$\Delta(w-p)$
1861–1913	1.00	0.20	4.2	1.09	0.80
1914–1920	14.6	14.0	1.5	−1.60	0.60
1921–1923	−12.2	−11.9	9.5	3.70	−0.30
1924–1938	0.50	−0.50	9.9	1.10	0.90
1939–1945	8.20	5.90	1.6	0.50	2.30
1946–1968	6.00	3.90	1.5	2.30	2.10
1969–1981	13.4	11.9	4.3	1.74	1.60
1982–2011	5.20	3.50	7.9	1.79	1.70

Table 6.1
Means of Δw, Δp, U_r, $\Delta(g-l)$ and $\Delta(w-p)$ over eight sub-periods in % p.a.

6.2.1 Step-indicator saturation

An extension of IIS to step-indicator saturation (SIS: see Castle, Doornik, Hendry, and Pretis, 2015) provides a more powerful approach to finding location shifts. Instead of defining a complete set of impulse indicators as in section 5.5, a complete set of increasing step indicators $\left\{1_{\{t\leq j\}}, j=1,\ldots,T\right\}$ is added. Step indicators are the cumulation of impulse indicators up to each next observation, as shown in the following Table.

$$\begin{array}{cc} \text{Impulses} & \text{Step shifts} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Table 6.2
IIS and SIS indicators

Figures 6.1 and 6.2 illustrate the 'split-half' approach of SIS applied to:

$$y_t = 15 + \lambda \times 1_{\{t\geq 23\}} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}[0,1] \tag{6.1}$$

where $\lambda = 0$ (null) then $\lambda = 5$. As with IIS, the three rows correspond to adding the first half of the indicators, dropping those, then adding the second half, then combining the selected indicators. The three columns report the indicators entered, the indicators retained, and the fitted and actual values of the selected model. When

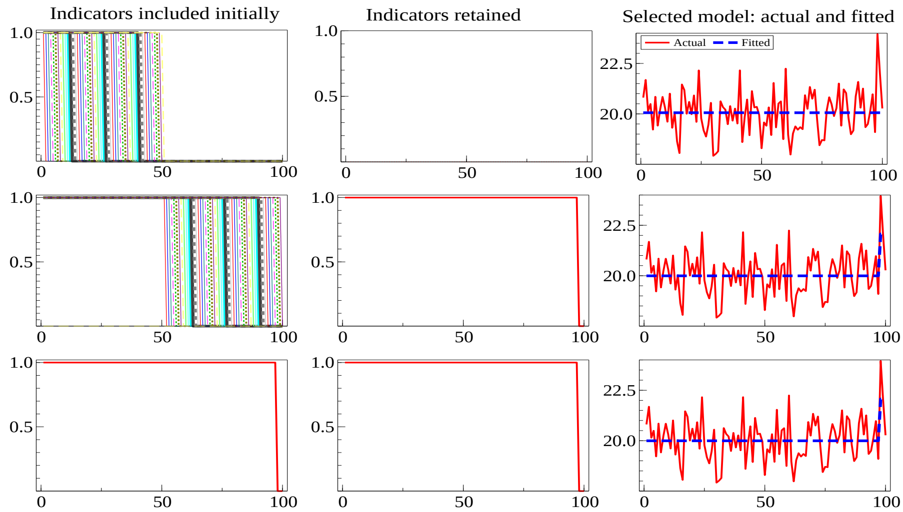


Figure 6.1
Illustrating SIS under the null of no shift in (6.1): $\lambda = 0$

$\lambda = 0$ and the first fifty step indicators are added, none is retained at 1% (row 1), but one second-half indicator is retained (row 2), so selecting over that retained indicator again keeps it.

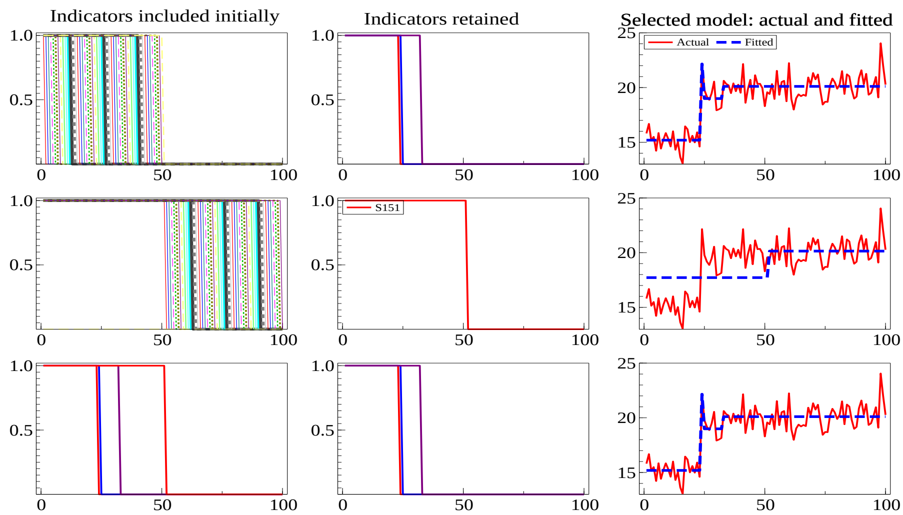


Figure 6.2
Illustrating SIS for a shift in (6.1): $\lambda = 5$

When $\lambda = 5$, three step indicators are retained at 1% from the first fifty (row 1), capturing the location shift and oscillations around it, and one retained from the second-half indicators (row 2), which is the one closest to the earlier location shift (a phenomenon that can be proved mathematically), and now selecting over the four retained indicators only keeps the first 3, as the second-half indicator is not needed to represent the location shift. The *Autometrics* version of SIS applied to this artificial data series just retains the step for the location shift. Like IIS, SIS can be used when selecting jointly with all the other modelling complications.

6.2.2 SIS in action

If we apply SIS to $\Delta(w-p)_t$, Δp_t, $\Delta(g-l)_t$, and $U_{r,t}$, we find similar but also somewhat different, outcomes to those in Table 6.1 as shown in Figure 6.3 where the indicators shown were selected at 0.1% significance, so are very unlikely to be due to chance.

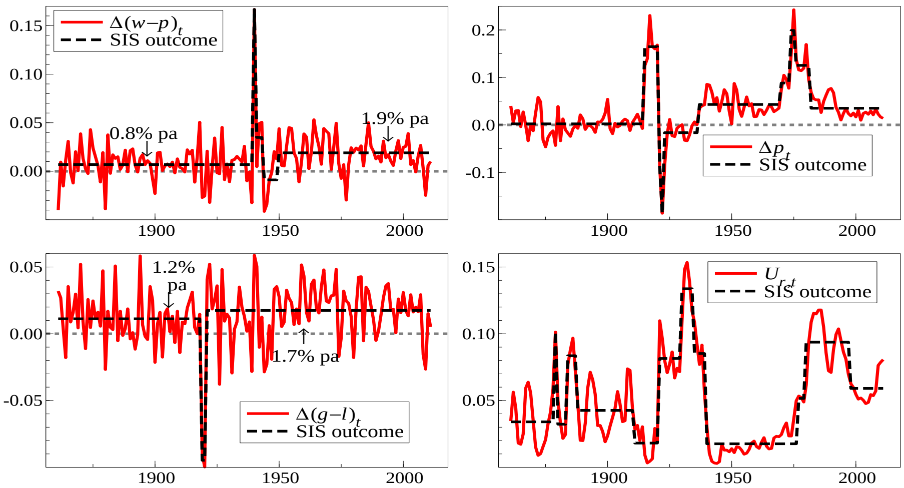


Figure 6.3
Location shifts in $\Delta(w-p)$, Δp, $\Delta(g-l)$ and U_r found by SIS.

The large mean shift is found for $\Delta(w-p)$ around 1945, and the spike in 1940, but with a small dip in between. The outcomes for Δp are as described earlier, but for $\Delta(g-l)$, its mean shift starts in 1921 following the very sharp drop over 1918-1919 as WWI ended (and the major flu' pandemic was under way). Thus, as Table 6.1 suggested, $\Delta(w-p)$ and $\Delta(g-l)$ do not co-break. Consequently, any model linking them will need to allow for their different shifts in addition to any direct connection. The picture changes again for U_r, where additional 'within regime' shifts are detected, with the business cycles of the early and mid 1880s being especially severe, and the impact of the Great Depression treated as a further shift. These extra shifts do not alter the earlier notion of four unemployment regimes.

6.3 Wage and price adjustments

Researchers have investigated many determinants of both real and nominal wage adjustment including unemployment; insiders versus outsiders; Trades Unions; worker-employer bargaining; staggered wage contracts; institutional factors; and price indexation among others for the former; and capacity utilization or the output 'gap'; the NAIRU; money growth; exchange rates; and terms of trade shocks for the latter (i.e., nominal wages). If nominal and real wage models both include Δp_t as a contemporaneous conditioning variable, they must be equivalent, so any other determinants should be the same. However, that would not apply if Δp_t was excluded from a nominal wage model, in which case the determinants of Δp_t would need to be included instead.

Some modellers focused on one basic determinant and one adjustment process, which has led to a proliferation of empirical claims. Since economic variables are highly inter-correlated and autocorrelated, as well as subject to both common and different location shifts, it is easy to imagine that some claims can achieve empirical 'corroboration' even when they could also be refuted. An obvious idea is to include in the GUM all the variables postulated as relevant in any of the main theories, and see which are actually relevant empirically when included with all the others.

6.3.1 Real wages and the impact of price inflation

Despite starting from a very large GUM, economic theory can still inform key aspects of a model's specification. In particular, to include all the potentially relevant variables assumes they operate equally over the data sample—but some variables may be dominant at some times, and irrelevant at others. Moreover, policy regimes change and that can alter which variables matter in different sub-periods.

As one example, when price inflation is low, there is little benefit to workers from seeking frequent pay negotiations (see Reis, 2006, for a related approach to 'inattentive producers'). However, if price inflation rises, it pays workers to be more attentive, and act to prevent erosion of their real wages. That reasoning suggests a non-linear reaction of real wages to inflation, where one possible mapping is:

$$\Delta(w - p)_t = f_t \Delta p_t + \cdots \tag{6.2}$$

where $f_t \simeq -1$ when $\Delta p_t \simeq 0$, so there is some real wage erosion, but at little cost to workers because inflation is low. Conversely, $f_t \simeq 0$ when Δp_t is large, as workers act strongly to avoid the substantial erosion that would otherwise occur. While Δp_t is always relevant because of (6.2), there would appear to be a less than complete, and possibly changing, impact of price inflation on real wages in a linear representation. Between these extremes of very low and very high price inflation there will be some erosion of real wages, but that too can be recouped by an equilibrium correction, albeit with a lag. Interestingly, one of the first EqCM models arose in the context of modelling real wages: see Sargan (1964).

The specific form for (6.2) used by Castle and Hendry (2009) was:

$$f_t = \frac{-1}{1 + \theta(\Delta p_t)^2}. \tag{6.3}$$

They set $\theta = 1000$ so that there was almost no erosion by $\Delta p_t = 0.1$ (approximately 10% p.a.). Figure 6.4 panel *a* shows the resulting response calculated from the historical data on Δp_t (panel *b* is discussed in Section 6.7).

Real wages have also adjusted almost completely by $\Delta p_t < -0.1$, presumably not because workers fought to achieve that, although employers certainly would, but in fact those data points were mainly determined by the indexation of wages during World War I continuing to operate after the war, as noted in Figure 2.10. Such a reaction seems bound to create the wage-price spirals that bedevilled the UK in the late 1960s and 1970s: once price inflation stimulates a wage reaction, firms' costs rise prompting further price increases.

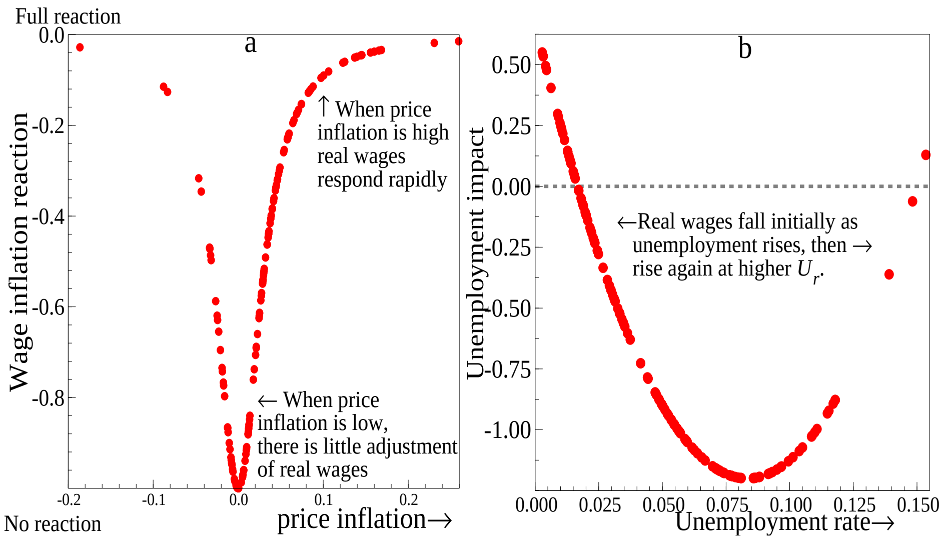


Figure 6.4
Non-linear responses (*a*) of real wages to inflation, and (*b*) to unemployment.

6.4 An empirical model of UK real wages

The initial GUM in Castle and Hendry (2009) for $\Delta(w-p)_t$ in (6.2) included 25 candidate variables to allow data evidence to determine which mattered, as well as the dynamics and the adjustment process (6.3). Selection of indicators using IIS at 0.1%, while retaining all the regressors without selection, found indicator variables for 1918, 1940, 1975+1977 (the Oil Crisis and a period with Incomes Policies), and a group for WWII, measuring 3%, 14%, −5%, and 3% respectively. Then the regressors in the GUM with these indicators were selected at 1%, yielding $\widehat{\sigma} = 1.24\%$ (of the level of real wages–see Section 1.2.1–a convention used below for equations explaining log variables), with no significant mis-specification tests.

$$\begin{aligned}
\Delta(w-p)_t &= \underset{(0.13)}{0.76}\ f_t \times \Delta p_t - \underset{(0.04)}{0.14}\ \Delta_2 U_{r,t-1} + \underset{(0.05)}{0.39}\ \Delta(g-l)_t \\
&+ \underset{(0.05)}{0.13}\ \Delta(g-l)_{t-2} - \underset{(0.01)}{0.08}\ \left(ulcp^* - \widehat{\mu}\right)_{t-2} + \underset{(0.002)}{0.010} \\
\mathsf{R}^2 &= 0.75 \quad \widehat{\sigma} = 1.24\% \quad T = 1863\text{–}2004 \qquad (6.4)
\end{aligned}$$

Only a few variables actually mattered as (6.4) records (indicators not shown).[1] The variable denoted $ulcp^*_t$—for real unit labour costs—is $(w-g+l-p)_t = c_t - p_t$ (where we came across c_t in Section 2.6.2), which is also the labour share, but adjusted for known changes in average hours worked. $\Delta_2 U_{r,t} = U_{r,t} - U_{r,t-2}$ is the change in unemployment over 2 years.

The coefficient of $f_t \Delta p_t$ is highly significant, and although somewhat less than unity, is not significantly so. Since that entails some real wage erosion, and only $0.39 + 0.13 \simeq 0.5$ of productivity changes are reflected in real wages within two years, equilibrium correction is needed to 'catch up' on past real wage losses. Real

[1] The indicators were estimated as $0.027 I_{1918} + 0.14 I_{1940} - 0.05(I_{1975} + I_{1977}) + 0.03\,(I_{1942} + I_{1943} - I_{1944} - I_{1945})$.

wages eventually converge to an equilibrium determined by $ulcp^* = \mu$, where μ is the long-run wage share, which is $\widehat{\mu} = 1.85$ over this period given the data units (prices and wages are indexes): see Figure 1.11, which shows the 'common trend' of real wages and productivity, and Figure 2.7 for a graph of the wage share (not adjusted for changes in hours). Unemployment only matters through changes, not its level, and those anyway just have a small effect. Despite the many variables initially allowed, the final model is relatively parsimonious and reasonably interpretable, although the intercept of 1% p.a. reveals unexplained real wage growth, so other variables, or location shift indicators, may matter, as we will investigate below.

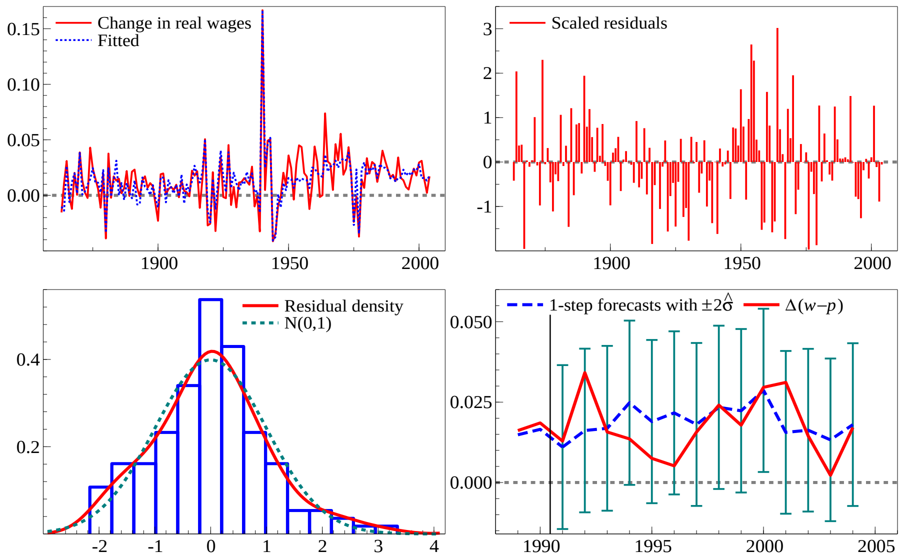
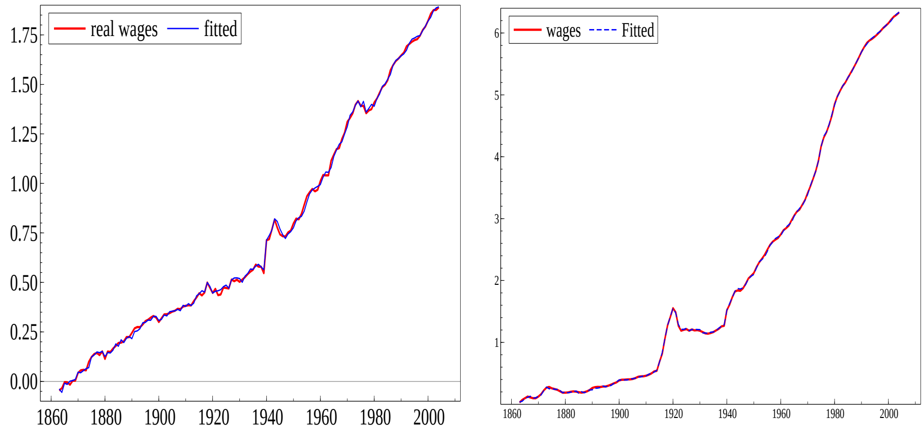


Figure 6.5
Graphical statistics of model fit and residual statistics for (6.4) and 1-step 'forecasts' with $\pm 2\widehat{\sigma}$ over 1991–2004

The graphs of the model fit, the residuals and their density are shown in Figure 6.5, together with some *ex post* 'forecasts' over 1991–2004 made using pre-1990 estimates. The fit is relatively poor during the post-war reconstruction 1950s and 60s, when unemployment was exceptionally low and real growth quite high, although the residuals are homoskedastic and near normal overall. Following the Thatcher era labour-market reforms, and the 1992 exit from the ERM with its large devaluation and expectations of subsequent higher inflation, constancy over the last period provides a useful check on the specification, which does not include any expectations measures often deemed important in wage models.

6.4.1 Graphical translation of real-wage inflation

It is possible to 'translate' the model of $\Delta(w-p)_t$ in (6.4) to explain either of the real-wage level $(w-p)_t$ given $(w-p)_{t-1}$, or the level of nominal wages w_t given p_t, w_{t-1}, p_{t-1}. The left panel below shows the former, and the right panel the latter.



Both show 'near perfect' fits, and would have R^2 values near unity, an illusion from the strong trends, as the residuals are the same as those in (6.4).

6.5 An expectations-augmented wage model

To test the possible importance of expectations about future wages in determining present wage, Δw_{t+1} was added to (6.4) using as instrumental variables $(R_S - R_L)_{t-i}$, $\Delta p_{o,t-i}$, $\Delta p_{w,t-i}$, Δp_{t-i} for $i = 1, 2$. These were selected as significant explanatory variables in the model for price inflation where $p_{w,t}$ and $p_{o,t}$ are world prices and oil prices respectively. The real wage was split into its nominal wage and price components, leading to:

$$\begin{aligned}
\Delta w_t \;&=\; -\underset{(0.10)}{0.15}\ \mathsf{E}_{t-1}[\Delta w_{t+1}] - \underset{(0.04)}{0.02}\ \Delta w_{t-1} + \underset{(0.003)}{0.014} + \underset{(0.10)}{1.1}\ \Delta p_t \\
&\quad + \underset{(0.17)}{0.70}\ f_t \times \Delta p_t + \underset{(0.08)}{0.46}\ \Delta(g-l)_t + \underset{(0.07)}{0.17}\ \Delta(g-l)_{t-2} \\
&\quad - \underset{(0.02)}{0.11}\ (ulcp^* - \widehat{\mu})_{t-2} - \underset{(0.06)}{0.15}\ \Delta_2 U_{r,t-1} \qquad (6.5) \\
\widehat{\sigma} \;&=\; 1.4\% \quad T = 1876 - 2001
\end{aligned}$$

where the same indicators were included but are not shown. The expectation is denoted $\mathsf{E}_{t-1}[\Delta w_{t+1}]$ and is equivalent to the forecast of Δw_{t+1} from the instruments listed above. As can be seen, its addition leaves the wage model almost unchanged, and that measure of expectations even has the 'wrong sign': similar findings for other models of inflation are reported in Castle, Doornik, Hendry, and Nymoen (2014), and in Chapter 8.

6.6 A new empirical model of UK real wages

Several developments prompted a new analysis of real-wage determinants, reported in Castle and Hendry (2014). First, seven new observations became available, spanning the turbulent period of the Great Recession, allowing a test of the general specification for continued constancy. We will consider how well the model 'forecasts' those in Section 6.8.2. Secondly, the test for non-linearity in Castle and Hendry (2010) suggested that more general non-linearities were required (Section 6.6.1). Thirdly, the extension of IIS (denoted SIS) was developed to address location shifts more directly (Section 6.2.1 above). Fourthly, there were several rival models, so encompassing tests thereof were feasible (section 6.6.2). Together, these led to a new empirical model of UK real wages in section 6.7. Finally, SIS also provided a powerful new test of super exogeneity (section 6.8.1).

6.6.1 Including other non-linear functions

The low-dimensional test for non-linearity in Castle and Hendry (2010) discussed in section 2.11 applied to (6.4) rejected linearity. However, their test had suggested a method for including general non-linear functions of variables during automatic model selection at the same time as IIS. Instead of estimating the linear model then checking for possible omitted non-linearities, add the powers and exponential functions of either the principal components or the current dated variables to the GUM. This led to including non-linear functions for $U_{r,t}$.

6.6.2 Encompassing

So far, the only encompassing test was in section 4.11.2 between the two misspecified equations. There is a range of possible tests, but the easiest to implement, and to understand, is that of nesting all the contending models in an 'artificial' GUM and testing the validity of reductions to each in turn. Moreover, even if none of the individual models proves valid, the most parsimonious yet acceptable reduction of that GUM provides a baseline for future studies.

The non-linear model of real wages in Nielsen (2009) had allowed for parameters shifting pre and post WWII, but encompassing tests against an updated version of (6.4) revealed that neither encompassed the other, nor did several other contenders considered by Castle and Hendry (2014), including a version of the logistic smooth transition model in section 2.11.

Combining all of these developments led to the model in the next section.

6.7 A SIS-based model of real wages

To allow for the many potentially relevant variables, the dynamic adjustments, the possible (and known) non-linearities, and the location shifts, the approach was as follows. First, the GUM included an intercept and the following regressors, all of which were specified as fixed, so not selected over initially, noting that perfect collinearities will be eliminated during later selection:

$\Delta(w-p)_{t-1}$, $\Delta(w-p)_{t-2}$,
$\Delta(g-l)_t$, $\Delta_2(g-l)_t$, $\Delta_2(g-l)_{t-1}$, $(\Delta(g-l)_t)^2$, $(\Delta(g-l)_{t-1})^2$,
$(\Delta(g-l)_t)\exp(-|(\Delta(g-l)_t)|)$, $(\Delta(g-l)_{t-1})\exp(-|(\Delta(g-l)_{t-1})|)$, $(\Delta(g-l)_t)^3$, $(\Delta(g-l)_{t-1})^3$,
$U_{r,t}$, $U_{r,t-1}$, $\Delta_2 U_{r,t}$, $\Delta_2 U_{r,t-1}$, $(U_{r,t}-0.05)^2$, $(U_{r,t-1}-0.05)^2$, $(U_{r,t}-0.05)^3$, $(U_{r,t-1}-0.05)^3$, $(U_{r,t}-0.05)\exp(-|(U_{r,t}|)$, $(U_{r,t-1}-0.05)\exp(-|(U_{r,t-1})|)$, $f_t\Delta p_t$, $f_{t-1}\Delta p_{t-1}$,
$\Delta^2 p_t$, $\Delta^2 p_{t-1}$, $\Delta^2 p_{t-2}$, $(\Delta p_t)^2$, $(\Delta p_{t-1})^2$, $(\Delta p_t)^3$, $(\Delta p_{t-1})^3$,

$(\Delta p_t)\exp(-|(\Delta p_t)|)$, $(\Delta p_{t-1})\exp(-|(\Delta p_{t-1})|)$,
$(w-p-g+l-\widehat{\mu})_{t-1}$, $(w-p-g+l-\widehat{\mu})_{t-2}$,
$(S_{2011}-S_{1946})\Delta u_{r,t}$, I_{1916}, I_{1977}, $(I_{1942}+I_{1943}-I_{1944}-I_{1945})$.

This specification builds on all the earlier empirical findings, so the formulation retains the non-linear reaction to price inflation in (6.3), and that the equilibrium correction is $(w-p-g+l-\widehat{\mu})$ (not corrected for hours) although the lag is selected. However, other non-linearities are of the general form in (2.17).

Secondly, SIS was applied at 0.1% so only the very largest outliers and shifts would be selected. This led to three step indicators being retained, for 1939, 1940 and 1943, implicitly entailing a step shift from 1944. Then the status of the fixed variables was removed, and the resulting equation was selected at 1%, leading to the model in (6.6).

Although the GUM differs from that in Castle and Hendry (2014), as well as using a different approach to selection, equation (6.6) is similar to that reported by them, with a slightly different dynamic structure. No diagnostic test rejects. The fit is considerably better than (6.4), where $\widehat{\sigma} = 0.0124$, yet variables in common have relatively similar coefficient estimates.

$$
\begin{aligned}
\Delta\left(w-p\right)_t = \underset{(0.003)}{0.031} &+ \underset{(0.041)}{0.303}\Delta\left(g-l\right)_t + \underset{(0.032)}{0.123}\Delta_2\left(g-l\right)_{t-1} + \underset{(0.012)}{0.725}\left(f_t\Delta p_t\right) \\
&- \underset{(0.028)}{0.158}\left(w-p-g+l-\widehat{\mu}\right)_{t-2} - \underset{(0.035)}{0.201}U_{r,t} - \underset{(0.045)}{0.14}\,\Delta_2 U_{r,t} \\
&+ \underset{(0.69)}{3.22}\left(U_{r,t}-\overline{U}_r\right)^2 - \underset{(0.029)}{0.145}\Delta^2 p_t - \underset{(0.011)}{0.149}S_{1939} + \underset{(0.014)}{0.166}S_{1940} \\
&- \underset{(0.008)}{0.044}S_{1943} - \underset{(0.008)}{0.025}(S_{2011}-S_{1946})\Delta u_{r,t} - \underset{(0.011)}{0.038}\,I_{1916} \\
&+ \underset{(0.007)}{0.050}\left(I_{1942}+I_{1943}-I_{1944}-I_{1945}\right) - \underset{(0.011)}{0.045}I_{1977} \qquad (6.6)
\end{aligned}
$$

$$
\begin{gathered}
\mathsf{R}^2 = 0.823;\ \ \widehat{\sigma} = 1.03\%;\ \ T = 1864-2004; \\
\chi^2_{\mathsf{nd}}(2) = 1.06;\ \ \mathsf{F}_{\mathsf{ar}}(2,123) = 0.06;\ \ \mathsf{F}_{\mathsf{arch}}(1,139) = 1.571; \\
\mathsf{F}_{\mathsf{het}}(21,116) = 0.92;\ \ \mathsf{F}_{\mathsf{reset}}(2,123) = 2.7;\ \ \mathsf{F}_{\mathsf{chow}}(7,125) = 1.03.
\end{gathered}
$$

To interpret the estimated equation, the short-run impact of $\Delta(g-l)$ is a little greater than 0.5 (note that $0.123\Delta_2(g-l)_{t-1}$ entails a response of 0.246 at annual rates), whereas the long-response is unity from the equilibrium correction, so the remaining shortfall is removed at about 16% p.a. by the EqCM term. Next, the reaction of real wages to price inflation erosion shown in Figure 6.4 *a* is a little smaller than unity at 73%, but no other term in Δp was retained including the other possible non-linear functions ($\Delta^2 p$ is the change in inflation). Thus, (6.3) seems to capture the impact of the level of inflation on real wages, and shows it depends on the magnitude workers face. Castle and Hendry (2014) show that the formulation is close to a logistic smooth transition.

The level of unemployment enters non-linearly: when expressed as a percentage (so 0.05 is 5%), the combined term can be written as $-.5U_{r,t}(1-6.4U_{r,t})$ and is shown in Figure 6.4 *b*. Thus, as unemployment increases, its negative impact on real wages increases till about 8% unemployment then starts to decrease, and even changes sign when unemployment exceeds about 15%. A possible explanation is that initially workers suffer a loss of bargaining power, but then movements down the marginal product of labour curve raise real wages for the more productive workers still employed. There is also a small negative impact from changes in unemployment. The regime-shift term in $\Delta \log U_{r,t}$ from Nielsen (2009) was also significant for the post WWII epoch, adding to the overall negative impact on real wages of increases in unemployment.

6.8 Evaluating the SIS-based wage model

Figure 6.6 records the actual and fitted values, the scaled residuals the residual density and the residual correlogram.

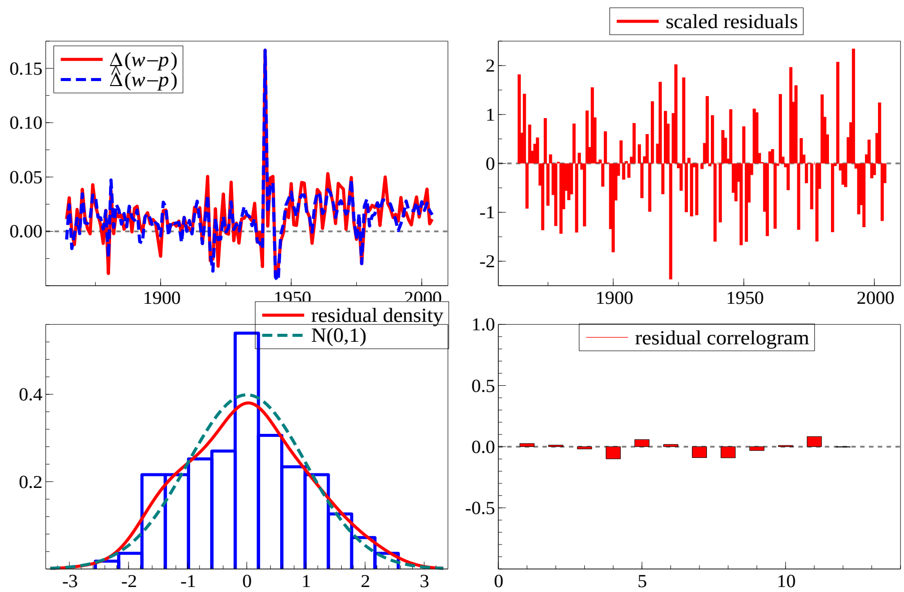


Figure 6.6
Graphical statistics for (6.6)

The graphs are consistent with the formal mis-specification tests, and now show a relatively constant fit over most of the 150 years. As (6.6) was designed to encompass previous empirical equations, there is no additional evaluation in that direction.

6.8.1 Testing super exogeneity

However, SIS can be used to test the super exogeneity of the k linear contemporaneous conditioning variables, denoted x_t say, in the selected model. Super exogeneity requires that shifts in the process generating x_t do not alter the parameters of the conditional model. *Autometrics* can be used to develop dynamic models for the x_t using SIS, either for each individual $x_{i,t}$ or for the vector x_t using a vector autoregression (VAR), although we only consider the former here (Castle and Hendry, 2014, used both IIS and SIS for a VAR).

Equations with up to 2 lags of $g-l$, Δp, U_r and $w-p$, but no contemporaneous variables, were formulated for the first three variables, and selected with SIS at $\alpha = 0.005$, then the retained indicators in the resulting marginal models were tested for significance in (6.6). Table 6.3 records for each marginal model (first column), how many step indicators were retained (q in the second column), the distribution of

Variable	q	null distribution	SIS test
$(g-l)_t$	2	F(2, 123)	0.77
Δp_t	7	F(7, 118)	1.87
$U_{r,t}$	14	F(14, 111)	1.37

Table 6.3
SIS based super-exogeneity test of (6.6).

the test statistic under the null (third column), and the value of the test statistic outcome (final column). In no case does the test reject, so the step shifts that were significantly perturbing the conditioning variables did not affect the model in (6.6), consistent with those regressors being super exogenous for its parameters.

6.8.2 'Forecasting' real wages

Task 20 Section 5.12 described the basic ideas behind 'forecasting' from a single equation when some of the regressors are contemporaneous.

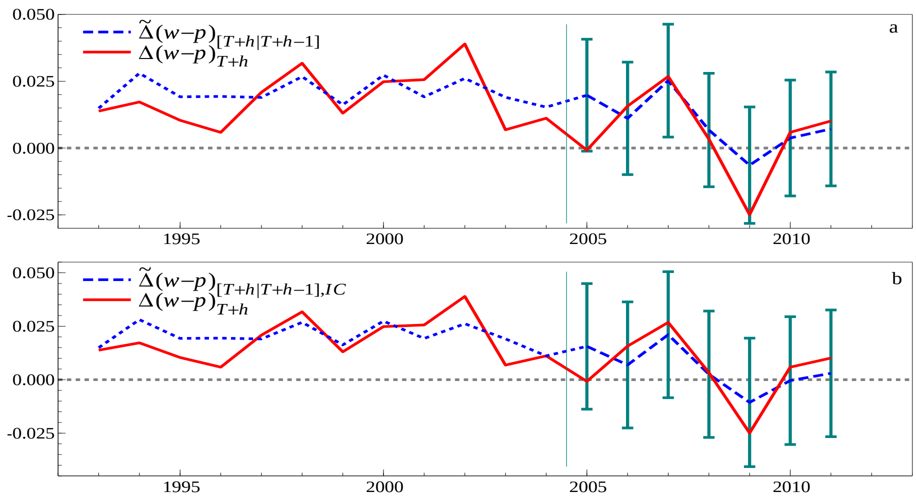


Figure 6.7
'Forecasting' real wages without and with intercept corrections

Figure 6.7 reports two sets of 1-step ahead forecasts from (6.6) for $\Delta(w-p)_t$ over 2005–2011. The top row (panel *a*) shows the forecasts made using the equivalent of (5.9), but with many more regressors. All the 'forecasts' lie inside the forecast intervals of $\pm 2\widehat{\sigma}_f$ calculated taking account of the parameter-estimation uncertainty, and the RMSFE = 0.0108, which is close to the in-sample $\widehat{\sigma}$ of 0.0103.

The bottom row (panel *b*) shows 'forecasts' made with an intercept correction (IC). An IC is an indicator variable with the value unity at and after the forecast origin, but zero before, so is almost equal to the last in-sample residual. Clements and Hendry (1998) show that ICs can offset systematic forecast errors made after a location shift, and Task 29 Section 8.8 analyses the properties of ICs based on an indicator variable for the final observation of the form used here. A seen in Figure 6.7, the IC improves the 'forecasts', leading to a RMSFE = 0.0098, which is smaller than $\widehat{\sigma}$, even though the IC indicator is estimated from just one observation. Both panels, however, show that there has not been a substantive shift in (6.6) over the Great Recession.

6.9 Chapter 6 key points

(A) Viable empirical models of real wages can be developed: the fundamental determinant of increased real wages since 1860 has been increased output per worker, sustained by increased capital, improved technology, and better education.
(B) The selected models of real-wage growth, $\Delta(w-p)_t$, can be mapped back to (w_t, p_t) to explain the nominal or real levels, given prices, $\{p_t\}$.
(C) The empirical analyses started from GUMs with all the theory-suggested variables, their dynamics, and non-linearities.
(D) Location shifts were reasonably captured by step-indicator saturation.
(E) $\Delta(w-p)_t$ had a non-linear response to Δp_t, showing increasing reactions to real-wage erosion from inflation.
(F) There was also a non-linear response to unemployment, consistent with *involuntary* unemployment as more workers are unemployed at higher real wages.
(G) Selecting location shifts by step indicator saturation did not preclude finding non-linearities, and vice versa, so both had roles to play.
(H) The long-run equilibrium was a constant wage share, which corrected past real-wage 'losses' from previous incomplete adjustments to changes in inflation, unemployment and productivity.
(I) Even with many candidate explanatory variables, lags and non-linear functions thereof, the relationship was perturbed by a few large shifts, mainly coinciding with wars.
(J) Despite selecting from more variables and indicators than observations, the final model was readily interpretable, passed a stringent super exogeneity test, and was constant over the Great Recession.

6.10 Task 21: How not to estimate wage models

First, estimate the regression of the log of wages, w, on that of prices, p, briefly discuss the resulting coefficient, and plot some graphs to help interpret how well the model describes w.

As before, click on the 'building block' Icon (fourth from the right on the Icon line, or Alt+**Y**). Choose `Models for time-series data`, and if not already set, select `Single-equation Dynamic Modelling using PcGive`.
Click on the `Formulate` button to bring up its dialog. Set `Lags` to zero, double click on 'w', and then on 'p' to formulate the equation, remembering that a Constant term is automatically added. `OK` brings up `Model Settings`, `OK` again to see the dialog for `Estimate`, set `Estimation ends` at 2011, and input 7 for `Less forecasts`

then OK to estimate the equation.

$$w_t = \underset{(0.016)}{0.235} + \underset{(0.009)}{1.38}\, p_t$$

$$\mathsf{R}^2 = 0.995 \;\; \widehat{\sigma} = 13.9\% \;\; \mathsf{F}_{\mathsf{ar}}(2,141) = 986.70^{**} \;\; \mathsf{F}_{\mathsf{arch}}(1,143) = 932.17^{**}$$

$$\chi^2_{\mathsf{nd}}(2) = 5.58 \;\; \mathsf{F}_{\mathsf{het}}(142) = 2.56 \;\; \mathsf{F}_{\mathsf{reset}}(2,141) = 24.75^{**} \;\; \mathsf{F}_{\mathsf{Chow}}(7,143) = 0.01$$

A coefficient of 1.38 does not make sense, as $w_t - p_t$ then rises or falls by 0.38% for every 1% that p_t rises or falls: it merely reflects the historical fact that wages have risen faster than prices on average. Despite $\mathsf{R}^2 = 0.995$, $\widehat{\sigma} = 13.9\%$ is huge, and Figure 6.8 reveals that the fit is very poor, the residuals systematic and non-normal, and there is massive residual autocorrelation. Almost all the mis-specification tests reject. Nevertheless, panel *a* in Figure 6.8 shows that the set of 1-step conditional 'forecasts', or more precisely, the out of sample fit, is fine, a result confirmed by the tiny outcome on the Chow test of parameter constancy. Bad models can still forecast well.

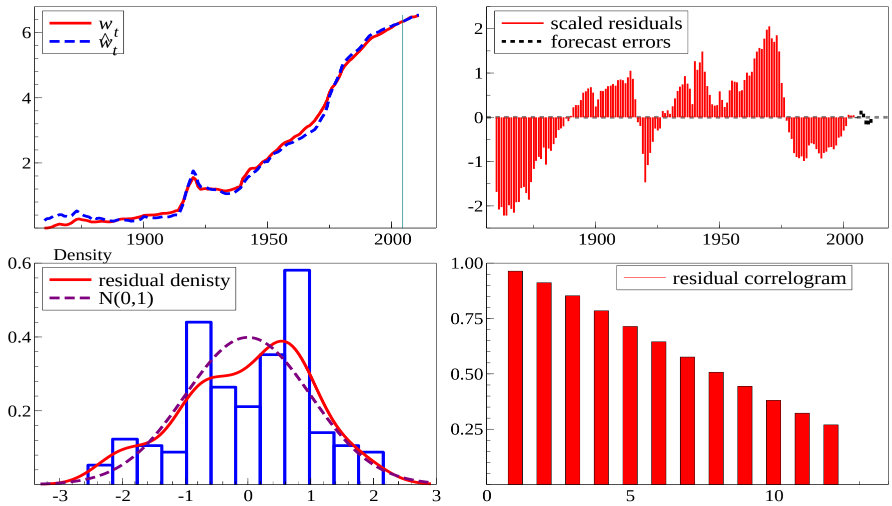


Figure 6.8
Graphical statistics for regressing w_t on p_t

Next, estimate the regression of wage inflation, Δw, on price inflation, Δp, and briefly discuss the resulting coefficient, again plotting graphs to help interpret how well the regression describes Δw. The *PcGive* commands are exactly as before, except you will need to clear the current model first, and instead select 'Dw' and 'Dp'.

$$\Delta w_t = \underset{(0.002)}{0.013} + \underset{(0.04)}{0.997}\, \Delta p_t$$

$$\mathsf{R}^2 = 0.85 \;\; \widehat{\sigma} = 2.36\% \;\; \mathsf{F}_{\mathsf{ar}}(2,140) = 0.14 \;\; \mathsf{F}_{\mathsf{arch}}(1,142) = 0.61$$

$$\chi^2_{\mathsf{nd}}(2) = 74.8^{**} \;\; \mathsf{F}_{\mathsf{het}}(2,141) = 0.82 \;\; \mathsf{F}_{\mathsf{reset}}(2,140) = 1.76 \;\; \mathsf{F}_{\mathsf{Chow}}(7,142) = 0.51$$

An almost 1-1 effect of price inflation on wage inflation is estimated, with $\widehat{\sigma} = 2.36\%$, so the fit is dramatically better despite a lower $\mathsf{R}^2 = 0.85$. Only the normality mis-specification test fails, mainly from the huge spike in real wages in 1940, as Figure 6.9 confirms.

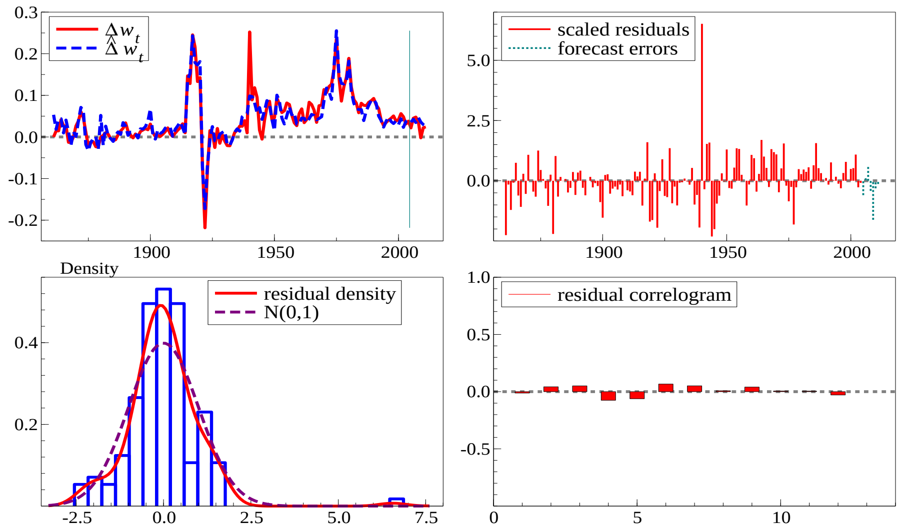


Figure 6.9
Graphical statistics for regressing w_t on p_t

Finally, estimate the regressions of wage inflation, 'Dw', on the unemployment rate, 'Ur', over the three sub-samples before 1914; between 1919 and 1939; and after 1945, and discuss your results. You should find the following:

$$\widehat{\Delta w_t} = \underset{(0.004)}{0.03} - \underset{(0.08)}{0.52}\ U_{r,t}$$

$$\mathsf{R}^2 = 0.45\ \ \widehat{\sigma} = 1.3\%\ \ \mathsf{F}_{\mathsf{ar}}(2,49) = 13.0^{**}\ \ \mathsf{F}_{\mathsf{arch}}(1,51) = 6.86^{*}$$
$$\chi^2_{\mathsf{nd}}(2) = 7.61^{*}\ \ \mathsf{F}_{\mathsf{het}}(2,50) = 2.83\ \ \mathsf{F}_{\mathsf{reset}}(2,49) = 5.93^{**}\ \ T = 1861-1913$$

$$\widehat{\Delta w_t} = \underset{(0.04)}{0.09} - \underset{(0.42)}{1.01}\ U_{r,t}$$

$$\mathsf{R}^2 = 0.235\ \ \widehat{\sigma} = 6.4\%\ \ \mathsf{F}_{\mathsf{ar}}(2,17) = 5.78^{*}\ \ \mathsf{F}_{\mathsf{arch}}(1,19) = 0.26$$
$$\chi^2_{\mathsf{nd}}(2) = 12.2^{**}\ \ \mathsf{F}_{\mathsf{het}}(2,18) = 0.03\ \ \mathsf{F}_{\mathsf{reset}}(2,17) = 3.24\ \ T = 1919-1939$$

$$\widehat{\Delta w_t} = \underset{(0.009)}{0.08} - \underset{(0.15)}{0.088}\ U_{r,t}$$

$$\mathsf{R}^2 = 0.008\ \ \widehat{\sigma} = 4.2\%\ \ \mathsf{F}_{\mathsf{ar}}(2,62) = 48.2^{*}\ \ \mathsf{F}_{\mathsf{arch}}(1,64) = 12.2^{*}$$
$$\chi^2_{\mathsf{nd}}(2) = 22.0^{**}\ \ \mathsf{F}_{\mathsf{het}}(2,,63) = 2.50\ \ \mathsf{F}_{\mathsf{reset}}(2,62) = 2.48\ T = 1946-2011$$

The parameter estimates and fits vary dramatically, so a single fit to the whole period will hardly describe the evidence. Omitting the war years loses the most informative periods of change, but is sensible if breaks are not also being handled.

6.11 Task 22: Step indicator saturation

The aim is to apply SIS for the observed variable $\Delta(w-p)_t$ using the tight level of 0.1% and compare the outcome to the mean shift shown in Figure 2.3. First, we will do a small analysis of the impact of a step indicator. Consider the equation:

$$y_t = \lambda 1_{\{t \geq T_1\}} + \epsilon_t \quad \text{for} \quad \epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2] \tag{6.7}$$

over $t = 1, 2, \ldots, T$ where $T_1 < T$, and $1_{\{t \geq T_1\}}$ is an indicator variable equal to zero when $t < T_1$ and unity when $t \geq T_1$. We will derive the formula for the least squares estimator $\widehat{\lambda}$ of λ, then interpret its impact on the residuals $\widehat{\epsilon}_t = y_t - \widehat{\lambda} 1_{\{t \geq T_1\}}$ from (6.7):

$$\begin{aligned} \widehat{\lambda} &= \sum_{t=1}^{T} y_t 1_{\{t \geq T_1\}} \Big/ \left(\sum_{t=1}^{T} 1_{\{t \geq T_1\}}^2 \right) \\ &= \sum_{t \geq T_1}^{T} y_t \Big/ \left(\sum_{t \geq T_1}^{T} 1 \right) = \frac{1}{T - T_1 + 1} \sum_{t \geq T_1}^{T} y_t = \overline{y}_{(2)} \end{aligned}$$

where $\overline{y}_{(2)}$ is the sub-sample mean of y_t over $T_1, T_1+1, \ldots, T$, using:

$$1_{\{t \geq T_1\}}^2 = 1_{\{t \geq T_1\}} = 1 \quad \text{for} \quad t \geq T_1.$$

Then:

$$\widehat{\epsilon}_t = y_t - \widehat{\lambda} 1_{\{t \geq T_1\}} = \begin{cases} y_t & \text{for } t < T_1 \\ y_t - \overline{y}_{(2)} & \text{for } t \geq T_1 \end{cases} \tag{6.8}$$

To implement SIS, select `Model`, `Formulate`, choose 'Dwp' with no lags, highlight the Constant, right click with the mouse and select `U: Unrestricted (fixed)`, `OK`, then tick the box for `Automatic model selection`, setting `Target size` to `Tiny:0.001` and double click on `Outlier and break detection`, setting that to `Step indicator saturation (SIS)`, `OK`. Set the sample to 1861–2011, and `OK` to see:

$$\begin{aligned} \widehat{\Delta(w-p)}_t &= -\underset{(0.018)}{0.16}\ \mathrm{S}_{1939} + \underset{(0.02)}{0.13}\ \mathrm{S}_{1940} + \underset{(0.013)}{0.044}\ \mathrm{S}_{1943} \\ &\quad - \underset{(0.008)}{0.028}\ \mathrm{S}_{1949} + \underset{(0.002)}{0.019} \\ \mathsf{R}^2 &= 0.43 \ \ \widehat{\sigma} = 1.78\% \ \ \mathsf{F}_{\mathsf{ar}}(2, 142) = 0.83 \ \ \mathsf{F}_{\mathsf{arch}}(1, 147) = 0.14 \\ \chi^2_{\mathsf{nd}}(2) &= 1.93 \ T = 1861 - 2011 \end{aligned}$$

After the search process is complete, Figure 6.10 shows the complete outcome you should see (after editing legends).

This SIS selection can check the main sub-sample mean shifts, originally found 'ocularly', by eliminating the indicators that tracked the WWII dip after the spike in 1940. How close were your estimates to those in Chapter 2 exercise *iv*), Section 2.18?

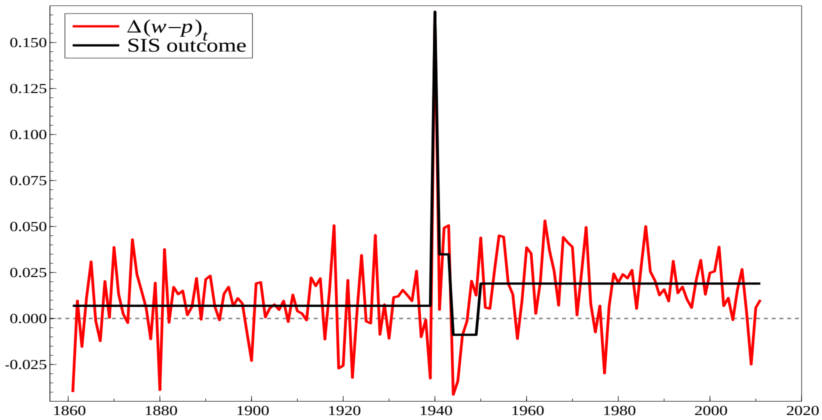


Figure 6.10
Mean shifts and outliers for real-wage growth

6.12 Task 23: Estimate a non-linear model

The approach builds on the test for non-linearity in Castle and Hendry (2010) (see section 2.11), but here uses the non-linear functions of the data variables rather than their principal components. To create these, select `Algebra` (Alt+**A**) and enter the following nine transforms:
UrSq=(Ur-0.05)ˆ 2;
DglSq = Dglˆ 2;
DpSq = Dpˆ 2;
UrCub=(Ur-0.05)ˆ 3;
DglCub = Dglˆ 3;
DpCub = Dpˆ 3;
UrExp=(Ur-0.05)*exp(-abs(Ur));
DglExp = (Dgl)*exp(-abs(Dgl));
DpExp = (Dp)*exp(-abs(Dp));
These can also be typed in the Results window, highlighted and computed by Ctrl+**A**.

Here we will only select from a simplified version of (6.6) using 1 lag. So `Model`, `Formulate`, set `Lags` to 1, and select 'Dwp', 'Dgl', 'Ur', 'wpglm', 'Dpfna', and the nine created non-linear terms (which will also enter lagged once), so there will be 28 variables (counting the dependent variable). `OK`, `OK` to estimate over 1864–2004. Very few of the coefficients will be individually significant. First we will test the significance of the 18 non-linear functions jointly. `Test`, tick `Exclusion Restrictions`, `OK`, and highlight the 18 non-linear terms, `OK` to see F(18,114) = 1.0853. Apparently there is no additional non-linearity. There are two reasons for that outcome: the first is the failure to jointly allow for location shifts, and the second is that one should only expect a few of the 18 non-linear terms to matter, so a joint F-test will lack power to reject.

6.13 Task 24: Model selection with SIS

The aim of the 2-stage selection process that follows is to use much tighter significance levels for shifts than variables. Continuing with the general non-linear equation, return to `Model`, `Formulate`, where the 28-variable equation should still be showing, highlight all 27 regressor variables, and right click with the mouse to select `U: Unrestricted (fixed)`, `OK`, then tick `Automatic model selection`. Set `Target size` to `Tiny:0.001` and double click on the option for `Outlier and break detection`, choosing `Step indicator saturation (SIS)`, `OK`, then set the sample to 1864–2004, with no forecasts and `OK`. SIS selects indicators for 1939, 1940, 1941, 1943, 1945, 1975, and 1977.

Return to `Formulate` to see the model with the step indicators, and `Clear status` on the previously fixed variables, then `OK`. Tick `Automatic model selection`, setting the `Target size` to `Small:0.01` and double click on `Outlier and break detection`, setting that to `None`, as we have already found the major shifts. Twenty terminal models 'look alike' when this overly simplified GUM is used. Somewhat tighter significance levels for non-linear than linear terms could also be used, clearing status for the 18 non-linear terms only but keeping the others as fixed, select, then clear all and re-select at `Medium:0.025`.

6.14 Chapter 6 exercises

1. Estimate the regression of $\Delta(w-p)_t$ on $\Delta(w-p)_{t-1}$ with a constant (both regressors marked as fixed), using SIS at 0.1%, over a sample ending in 2011, with 7 observations retained for forecasting. Discuss the coefficient estimates and the model evaluation statistics.
2. Can you explain why the outcome is close to that found in Task 22, Section 6.11?
3. How well does the forecast RMSE compare to that of the model used in Section 6.8.2? Is this a fair comparison, as the forecasts from *i*) are genuinely *ex ante*?
4. What does the fact that in both cases the forecast RMSEs are close to their respective in-sample $\widehat{\sigma}$s tell us about forecasting real wages over the Great Recession?
5. Prove that the 1-step forecasts for a variable Δy_t are the same as those for y_t when y_{t-1} is a regressor. Prove that such a claim is false when when y_{t-1} is not a regressor. Consequently, discuss the distinction between transforming the dependent variable in an equation to a difference by subtracting its lagged value, and differencing the variables in an equation.

Chapter 7
Modelling UK money demand

Chapter 7 guide posts

1. As with other aspects of the macroeconomy, there have been huge changes to the financial system since 1860: Section 7.1.
2. To create a measure of money relevant throughout, namely the broad component M4, various time series had to be spliced together: Section 7.2.
3. An analysis of motives for holding money guides the formulation of the GUM: Section 7.3 .
4. An earlier selected empirical model of money demand matches a theory with nominal short-run adjustment within bands and real long-run reactions shifting those bands in line with inflation and real income growth: Section 7.4.
5. The long-run equilibrium of the model is a relation between the inverse velocity of circulation of money (namely the log ratio of real money to real GDP) and the net interest rate cost from holding money, and updating that earlier model after recent data revisions delivers similar results: Section 7.5.
6. Re-estimating the money-demand model over the longer sample period to 2011 produces recognizably similar coefficients to the initial equation, and the cointegrating relation found up to 1975 is almost unchanged: Section 7.6.
7. However, step-indicator saturation is needed to capture several important location shifts since 1970: Section 7.7.

7.1 Money demand and supply

As we have seen in earlier chapters, a household in 1860 would have had little money, receiving it as a weekly wage, and spending almost all, including repaying short-term loans, before the next payday. Money was a flow to most people: only the wealthy had bank accounts, holding a stock of money as an asset, or as a precaution against negative events. When economists discuss the demand for money, all three reasons—transactions, asset and precautionary—occur, noting that the outstanding stock has to be held by someone at every point in time.

Historically, inflation was seen as the process whereby prices of goods and services rose in terms of the amount of commodity money (gold or silver) that had to be exchanged to purchase them. This was essentially a fall in the price of 'money', usually occasioned by an increase in its volume (see the elegant discussion of the inflation of the 16th Century from the influx of gold from South America to Europe by Hume, 1752). Deflation occurred when the supply of goods rose (as after the industrial revolution) without a concomitant increase in the supply of gold or silver, making 'money' relatively scarcer. In both cases, supply-side forces drove the outcome, either from an 'exogenous' increase in the supply of precious metals (as with the Californian gold rush of 1848, followed by the Australian in 1851, as well as many others), or conversely from prolific harvests or technological improvements.

The introduction of a large number of financial innovations over time, from bills of exchange, paper money and fractional-reserve banking, through personal cheque books in the 1830s, the telegraph in the 1860s, credit cards, to ATMs and on-line banking, as well as the creation of many different types of asset that can count as 'money', have complicated that story (already seen to be a problem by Marshall, 1926). At the start of our period, Building Societies (equivalent to Savings and Loans in the USA), were small cooperative institutions. They grew greatly over the next century, and began to rival commercial banks in their scale. Indeed, throughout the 1990s, most large Building Societies converted to commercial banks and commenced borrowing on wholesale money markets, at a time when bank liquidity ratios were allowed to decrease, setting the scene for the financial crisis where many previous Societies went into liquidation or were taken over. Changes in the spectrum of competing interest rates from financial innovation have plagued previous econometric models of various measures of money both broad (now measured by M4) and narrow, such as M1 (notes and coin plus demand deposits at commercial banks), sometimes called transactions money: see Hacche (1974) and Hendry and Mizon (1978); Coghlan (1978) and Hendry (1979); as well as Goldfeld (1976) and Baba, Hendry, and Starr (1992) for the USA.

Moreover, the roles of Central Banks have changed greatly from 'controlling the money supply' under the Gold Standard that operated till the 1930s, through a plethora of regimes, to now targeting inflation and unemployment. In the UK, money creation is akin to that of virtual particles in physics, where an asset and a matching liability are simultaneously produced (or annihilated), with recourse to discounting assets at the Central Bank if liquidity becomes scarce. In effect, the quantity of money outstanding is usually determined by the demand to hold the stock of it, although the Government could have an impact through unfunded budget deficits, but even the massive £375 billion purchases of Government debt by the Bank of England was accompanied by a fall in the volume of broad money.

Thus, the demand for the assets that can count as 'money' has become the main focus, although there remains a tradition (exemplified by Friedman and Schwartz; Friedman and Schwartz, 1963, 1982, among others) that inflation is primarily a monetary phenomenon, criticized by Hendry and Ericsson (1991). Friedman and Schwartz (1982) estimate demand for money functions with real money as the dependent variable, then solve them to express prices as dependent on nominal money, which is treated as 'exogenous'. We will address that theory by checking the validity of 'inverting' the status of money and prices in a model of UK money demand in Task 25 Section 7.9.

7.2 Money and interest rates

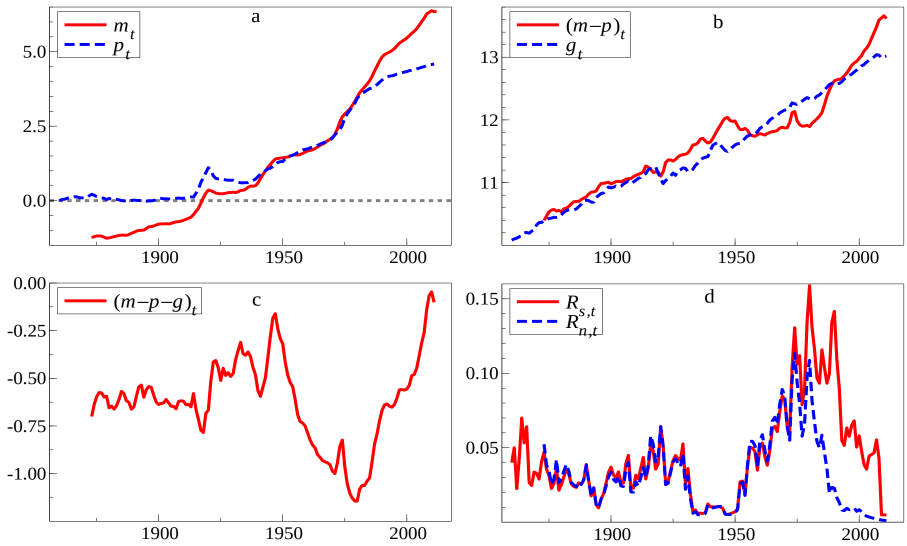


Figure 7.1
Time series of m_t, $(m-p)_t$, $(m-g-p)_t$, $R_{n,t}$ and $R_{S,t}$

Figure 7.1 panel *a* records the time series on the log of nominal broad money (created by splicing ever-increasing coverage of the data series on 'money' over time, discussed in the Appendix) with the log of prices. Panel *b* plots the logs of the stock of real money (in constant prices) and real GDP, which shows major departures between them. Real money increased markedly during the Thatcher 'monetary control' regime, which also coincided with substantial deregulation of the banking system, then slowed during the recession of the late 1980s–early 1990s before rising sharply again. Panel *c* plots the log of the relative amount of money in circulation in the UK compared to nominal GDP, which rose almost 3-fold from its low 1980, ending near its highest value, as discussed in Section 1.12. Finally, panel *d* records the short-term interest rate (on 3-month Treasury bills) $R_{S,t}$, and the competing interest rate net of the own interest rate, denoted $R_{n,t}$. These are essential the same until 1985, after which they depart markedly, with the latter falling to near zero by the late 1990s as the definition of money has become broader with a larger proportion of 'money' paying interest: see e.g., Ericsson, Hendry, and Prestwich (1998).

Figure 7.2 panel *a* plots Δp_t with its mean shifts (section 2.8 defined the eight sub-periods); panel *b* shows the time series plots of Δp_t and Δm_t; panel *c* records the corresponding location shifts in Δm_t; and panel *d* the scatter plot of Δp_t against Δm_t showing the dates and a regression line with projections. Panel *b* reveals many systematic departures between Δp_t and Δm_t with an absence of co-breaking, highlighted in panel *c* by the different location shifts in Δm_t found by SIS. Finally, the upward slope of the regression in panel *d* cannot be interpreted as 'money causes inflation' because as prices rise, people need more money to buy the same quantities of goods and services. Indeed, not only do m_t and p_t fail to cointegrate (even

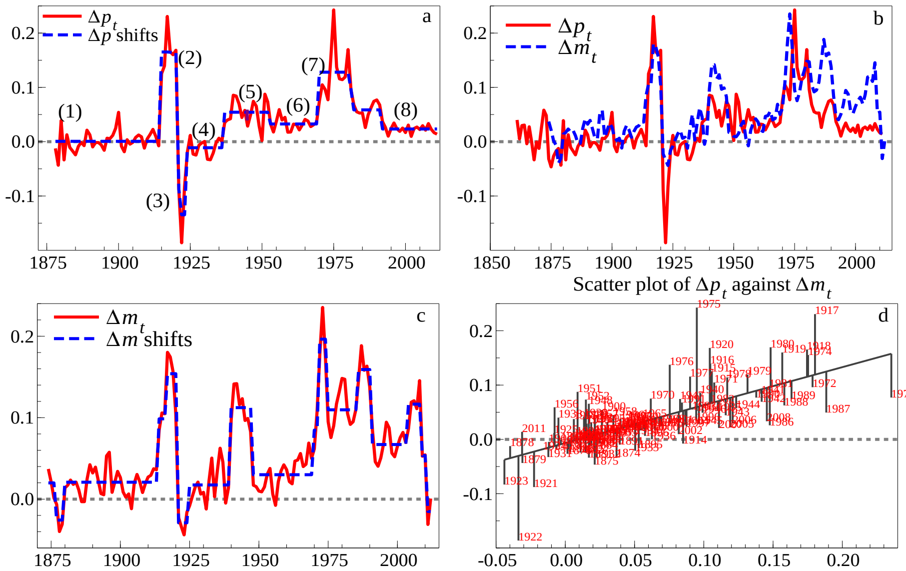


Figure 7.2
Location shifts in price inflation and money growth

allowing for shifts), using only Δm_t to 'explain' Δp_t (with 4 lags of each, first selecting by SIS at 0.1%, then over regressors at 1%) yields the long-run relation $\Delta p = 0.39\Delta m$ which is well short of proportionality with $\widehat{\sigma} = 0.021$. In comparison, an autoregression in Δp_t selected by SIS has $\widetilde{\sigma} = 0.022$. Money may be one of the determinants of price inflation, but is clearly not the only one: Chapter 8 will examine that issue. Here, we investigate the demand for broad money.

7.3 Formulating a demand for money model

Why would anyone want to hold paper money? It has no intrinsic value (unlike commodity money), is easily eroded by inflation, and its use for transacting depends on the willingness of others to accept it in exchange for real goods and services. The answer given by Starr (2012) is to pay taxes, as Governments both issue money and agree to accept it back for that purpose. Once it has such a use, other transactions follow suit.

Having given money the status of 'legal tender' the next question is how much money would economic agents then wish to hold? There are a plethora of theories of money demand, although many are only relevant to 'narrow' or transactions demand, summarized in Ericsson, Hendry, and Prestwich (1998). Various motives have been proposed: to finance 'physical' transactions, to facilitate speculation in financial markets, as a store of value, and as a precaution against short-term fluctuations in incomes. All seem likely to depend on levels of expenditure or income (and probably wealth), and possibly on the opportunity cost of holding money as an idle asset. That cost is the interest it could have earned when invested in other safe but interest-bearing assets (like 3-month Treasury Bills), and the depreciation in value over time of all nominal assets from inflation.

These ideas lead to a general formulation of aggregate money demand (denoted m^d in logs) depending on GDP, g, prices, p, and the opportunity cost, often measured by the interest rate on an outside asset, R_S less the interest rate paid on the proportion of M that earns interest R_o (like deposit accounts) denoted R_n, and inflation Δp:

$$m^d = f\left(g, p, R_n, \Delta p\right) \tag{7.1}$$

or in a log-linear form:

$$m^d = \theta_1 g + \theta_2 p + \theta_3 R_n + \theta_4 \Delta p \tag{7.2}$$

where $\theta_1 > 0$, $\theta_2 > 0$, $\theta_3 < 0$, $\theta_4 < 0$, often with $\theta_1 = \theta_2 = 1$ so real money demand is proportional to real GDP. A specification like (7.2) is an equilibrium relation, so at best corresponds to a cointegrating relationship, as a model of actual m_t would need to allow for dynamic adjustments, possible non-linearities, and shifts.

7.4 Modelling the demand for money

For the period 1878–1975 then available, Hendry and Ericsson (1991) found that the inverse velocity of circulation of money, namely $(m - g - p)_t$, cointegrated with the competing interest rate $R_{S,t}$ (which coincided with the net interest rate $R_{n,t}$ over most of their sample), as:

$$\widetilde{u}_t = \left(m - p - g\right)_t + 0.309 + 7R_{S,t} \tag{7.3}$$

where $\{\widetilde{u}_t\}$ did not have a unit root.

Their selected model, which included a non-linear equilibrium correction in $\widetilde{u}$, denoted by $e_{t-1} = (\widetilde{u}_{t-1} - 0.2)\widetilde{u}^2_{t-1}$, was:

$$\begin{aligned}
\Delta(m-p)_t \;=\;& \underset{(0.06)}{0.47}\,\Delta(m-p)_{t-1} - \underset{(0.04)}{0.11}\,\Delta^2(m-p)_{t-2} - \underset{(0.04)}{0.59}\,\Delta p_t \\
&+ \underset{(0.05)}{0.41}\,\Delta p_{t-1} - \underset{(0.006)}{0.017}\,\Delta r_{S,t} - \underset{(0.019)}{0.078}\,\Delta_2 r_{L,t} \\
&- \underset{(0.19)}{1.15}\,\widetilde{e}_{t-1} + \underset{(0.006)}{0.034}\,(D_1 + D_3) + \underset{(0.002)}{0.007} \\
&+ \underset{(0.010)}{0.071}\,D_4 + \underset{(0.020)}{0.090}\,D_4 \Delta r_{S,t}
\end{aligned} \tag{7.4}$$

$$\begin{aligned}
\mathsf{R}^2 &= 0.88 \quad \widehat{\sigma} = 1.48\% \quad \mathsf{F}_{\mathsf{ar}}(6,81) = 1.19 \quad \mathsf{F}_{\mathsf{arch}}(2,83) = 1.31 \\
\chi^2_{\mathsf{nd}}(2) &= 2.9 \quad \mathsf{F}_{\mathsf{het}}(18,68) = 0.43 \quad T = 1878 - 1975
\end{aligned}$$

where $D_1 + D_3$ were dummy variables taking the value unity during the two world wars respectively and zero elsewhere, and D_4 is a dummy which is unity only over the period 1971–75. Changes to financial regulations (called Competition and Credit Control) over 1971–75 required changing R_S to R_n and adding both the dummy D_4 and its interaction with changes in interest rates, $D_4\Delta r_{S,t}$. The mis-specification tests show that (7.4) was congruent, and they demonstrated that it was constant over the sample to 1970, but that 'inverting' to have Δp_t as the dependent variable with Δm_t treated as exogenous was not constant. The small net negative coefficient of inflation in (7.4), obtained by combining the coefficients of Δp_t and Δp_{t-1}, shows some erosion of the value of real money holdings.

To interpret (7.4), the formulation of e_{t-1} preserves the sign of the disequilibrium feedback, as the cubic component dominates, but there are two possible equilibria, one when $u = 0$, the usual cointegration solution, and the other at $u = 0.2$. The second seems mainly to account for the discrepancy between $m - p$ and g over 1921–1955 that is visible in Figure 7.1 panel *b*, although e_{t-1} operates over the whole sample period. Real money growth is strongly autoregressive, and inflation and changes in (log) interest rates have net negative effects as expected.

Alternatively, (7.4) can be expressed in terms of nominal short-run money demand:

$$\Delta m_t \approx 0.41\Delta p_t + 0.11\Delta^2 p_{t-2} + 0.47\Delta m_{t-1} - 0.11\Delta^2 m_{t-2} \cdots$$

with long-run demand being in real terms as determined by the cointegrating relation (7.3). That matches a theory of money demand where in the short-run, agents adjust their holdings within bands, and in the longer run adjust those bands as the price level and income change: see Miller and Orr (1966) and Milbourne (1983). Finally, money demand increased by about 4% during both World Wars.

7.5 Updating the demand for money model

Although there have been important data revisions since their study, it was possible to obtain a reasonably close replication of (7.4):

$$\begin{aligned}
\Delta(m-p)_t \;=\;& \underset{(0.07)}{0.40}\,\Delta(m-p)_{t-1} - \underset{(0.05)}{0.14}\,\Delta^2(m-p)_{t-2} - \underset{(0.05)}{0.62}\,\Delta p_t \\
&+ \underset{(0.06)}{0.38}\,\Delta p_{t-1} - \underset{(0.006)}{0.025}\,\Delta r_{n,t} - \underset{(0.031)}{0.104}\,\Delta_2 r_{L,t-1} \\
&- \underset{(0.54)}{2.80}\,\widetilde{e}_{t-1} + \underset{(0.006)}{0.037}\,(D_1 + D_3) + \underset{(0.002)}{0.009} \\
&+ \underset{(0.010)}{0.065}\,D_4 + \underset{(0.028)}{0.087}\,D_4\Delta r_{n,t}
\end{aligned} \tag{7.5}$$

$$\begin{aligned}
\mathsf{R}^2 \;=\;& 0.87\;\; \widehat{\sigma} = 1.62\% \;\; \mathsf{F}_{\mathsf{ar}}(2,85) = 1.66 \;\; \mathsf{F}_{\mathsf{arch}}(1,96) = 0.004 \\
\chi^2_{\mathsf{nd}}(2) \;=\;& 7.16^* \;\; \mathsf{F}_{\mathsf{reset}}(2,85) = 0.69 \;\; \mathsf{F}_{\mathsf{het}}(18,79) = 1.50 \;\; T = 1878 - 1975
\end{aligned}$$

Most of the coefficients in (7.5) in common with (7.4) have the same signs and similar magnitudes, noting that $r_{n,t}$ replaces $r_{S,t}$, and one lag length has changed, although the coefficient of e_{t-1} is larger. The additional variable Δg_t had been insignificant in (7.4), but was significant if added to (7.5). As long-run money demand depends on g, it had been a surprise that its change (and lags thereof) was insignificant when included in (7.4).

7.6 Extending the demand for money model

There have been a number of updates of equations like (7.4) by other authors, including Escribano (2004) who examines a range of alternative non-linear equilibrium correction formulations, and concludes with a constant-parameter representation to 2000 similar to (7.4). The long-run solution in Hendry (2001) which was estimated as a cointegrating relation over 1870–1991 holds essentially unchanged to 2011 when estimated as a static regression, despite continuing major changes to the UK's financial system:

$$\widetilde{u}_t = (m - p - g)_t + 0.38 + 7.4R_{n,t} \tag{7.6}$$

where $R_{n,t}$ is the opportunity cost of holding money. Consequently, the previous non-linear equilibrium correction term e_{t-1} was retained.

Equation (7.7) records the outcome of extending the model including Δg_t over the additional quarter of a century 1976 to 2011, so the whole sample is $T = 1878-2011$:

$$\begin{aligned}\Delta(m-p)_t = \; & \underset{(0.07)}{0.60}\,\Delta(m-p)_{t-1} - \underset{(0.05)}{0.10}\,\Delta^2(m-p)_{t-2} - \underset{(0.06)}{0.71}\,\Delta p_t \\ & - \underset{(0.07)}{0.51}\,\Delta p_{t-1} - \underset{(0.007)}{0.021}\,\Delta r_{n,t} - \underset{(0.027)}{0.044}\,\Delta r_{L,t-1} \\ & - \underset{(0.25)}{1.26}\,e_{t-1} + \underset{(0.007)}{0.032}\,(D_1 + D_3) + \underset{(0.003)}{0.003} \\ & + \underset{(0.001)}{0.039}\,D_4 + \underset{(0.023)}{0.055}\,D_4\Delta r_{S,t} + \underset{(0.07)}{0.19}\,\Delta g_t\end{aligned} \tag{7.7}$$

$$\begin{aligned}\mathsf{R}^2 &= 0.77 \;\; \widehat{\sigma} = 2.12\% \;\; \mathsf{F}_{\mathsf{ar}}(2,120) = 1.46 \;\; \mathsf{F}_{\mathsf{arch}}(1,132) = 2.61 \\ \chi^2_{\mathsf{nd}}(2) &= 4.79 \;\; \mathsf{F}_{\mathsf{het}}(20,113) = 2.08^{**} \;\; \mathsf{F}_{\mathsf{reset}}(2,120) = 0.72\end{aligned}$$

The estimated coefficients are recognizably similar, especially compared to those in (7.4), but the fit is much poorer and the heteroskedasticity mis-specification test rejects.

7.7 Recent location shifts in the demand for money model

Both of these problems occur primarily because of the recent turbulent period, and given the step shifts seen in Figure 7.2c and the marked departure between the time-series behaviour of Δm_t and Δp_t after 1980, suggests applying step-indicator saturation, selecting at $\alpha = 0.001$ so only the most important shifts are found.

$$\begin{aligned}\Delta(m-p)_t = \; & \underset{(0.06)}{0.57}\,\Delta(m-p)_{t-1} - \underset{(0.04)}{0.12}\,\Delta^2(m-p)_{t-2} - \underset{(0.050)}{0.67}\,\Delta p_t \\ & + \underset{(0.06)}{0.51}\,\Delta p_{t-1} - \underset{(0.006)}{0.021}\,\Delta r_{n,t} - \underset{(0.024)}{0.060}\,\Delta r_{L,t-1} \\ & - \underset{(0.26)}{1.56}\,e_{t-1} + \underset{(0.007)}{0.030}\,(D_1 + D_3) - \underset{(0.012)}{0.020} + \underset{(0.008)}{0.051}\,D_4 \\ & + \underset{(0.02)}{0.05}\,D_4\Delta r_{n,t} + \underset{(0.06)}{0.17}\,\Delta g_t + \underset{(0.010)}{0.03}\,\mathsf{S}_{1973} - \underset{(0.015)}{0.06}\,\mathsf{S}_{1979} \\ & + \underset{(0.015)}{0.05}\,\mathsf{S}_{1981} - \underset{(0.008)}{0.04}\,\mathsf{S}_{1999} + \underset{(0.013)}{0.05}\,\mathsf{S}_{2008}\end{aligned} \tag{7.8}$$

$$\begin{aligned}\mathsf{R}^2 &= 0.86 \;\; \widehat{\sigma} = 1.81\% \;\; \mathsf{F}_{\mathsf{ar}}(2,115) = 1.56 \;\; \mathsf{F}_{\mathsf{arch}}(1,132) = 0.12 \\ \chi^2_{\mathsf{nd}}(2) &= 2.15 \;\; \mathsf{F}_{\mathsf{het}}(25,108) = 1.12 \;\; \mathsf{F}_{\mathsf{reset}}(2,115) = 0.46\end{aligned}$$

Although five step shifts are retained, the other coefficient estimates are not greatly altered. The diagnostic tests are all insignificant, though the fit remains less good than over the shorter period to 1975.

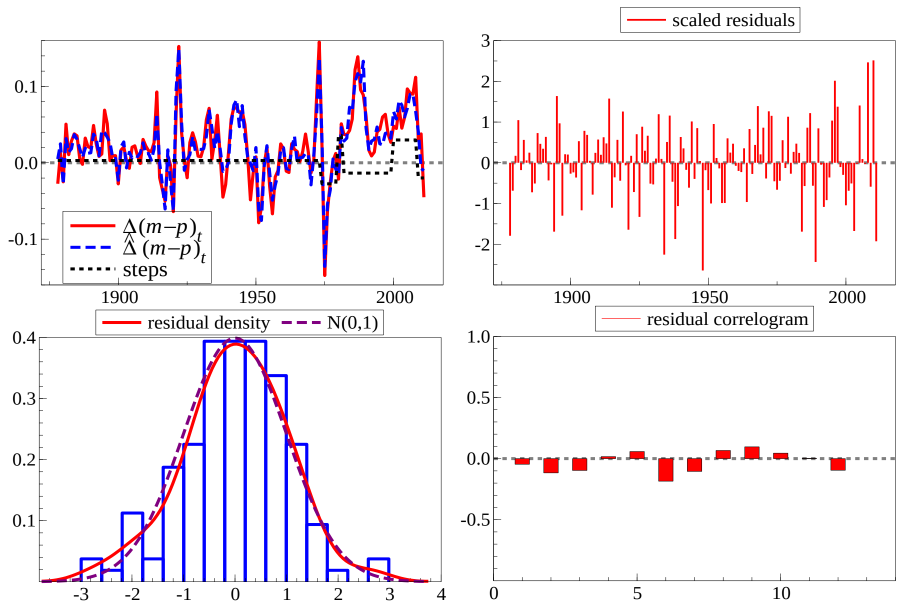


Figure 7.3
Money model extension using SIS

Figure 7.3 shows the graphical outcome, where panel *a* reports the combination of the step indicators. The main periods not captured by the economic variables are 1980–1999, where a negative location shift was needed, followed by a large positive step over 2000–2008.

7.8 Chapter 7 key points

(A) There have been huge changes in the financial system over the last 150 years.
(B) This necessitates creating a spliced time series to measure money, where the broad component M4 is used here.
(C) Updating the earlier model in Hendry and Ericsson (1991) for data revisions delivered similar results to theirs.
(D) The empirical models reasonably matched a theory of money demand with nominal short-run adjustment within bands and real long-run reactions shifting those bands in line with inflation and real income growth.
(E) The cointegrating relation was almost unchanged when extending the sample period from 1976 to 2011.
(F) Re-estimating the dynamic money demand model over the longer sample revealed recognizably similar coefficients to the initial equation.
(G) However, step-indicator saturation was needed to capture several important location shifts since 1970.

7.9 Task 25: 'Inverting' money-demand equations to explain inflation

As noted in Section 7.1, Friedman and Schwartz (1982) solve equations with real money to express prices as the dependent variable, with nominal money treated as 'exogenous'. Because (7.8) has a dependent variable $\Delta(m-p)_t$, it can be expressed either with Δm_t or Δp_t on the left-hand side with the other component carried to the right-hand side. As Δp_t is already a regressor, (7.8) is equivariant under this change, other than the coefficient of Δp_t changing. However, collecting Δp_t terms on the left and moving Δm_t as an 'exogenous' variable to the right can alter estimates radically, as the direction of conditioning matters considerably.

To conduct this Task, first formulate and estimate (7.7), mark all regressors as `fixed`, and re-estimate, ticking `Automatic model selection`, using `Step-indicator saturation` at a `Tiny:0.001` `Target size` to obtain (7.8). Now return to `Formulate`, and delete Dmp as the dependent variable, mark Dp as `Y:endogenous`, and add Dm and Dm_1 as regressors, removing Dmp_1. (7.9) shows the estimates of this inflation model.

$$\begin{aligned}
\Delta p_t \;=\;& \underset{(0.08)}{0.35}\ \Delta p_{t-1} + \underset{(0.12)}{0.79}\ \Delta m_t + \underset{(0.04)}{0.09}\ \Delta r_{L,t-1} - \underset{(0.13)}{0.24}\ \Delta m_{t-1} \\
&+ \underset{(0.44)}{1.61}\ e_{t-1} + \underset{(0.01)}{0.04}\ \Delta r_{n,t} + \underset{(0.07)}{0.11}\ \Delta^2(m-p)_{t-2} - \underset{(0.10)}{0.04}\ \Delta g_t \\
&- \underset{(0.01)}{0.00}\ (D_1+D_3) - \underset{(0.02)}{0.03}\ D_4 - \underset{(0.03)}{0.08}\ D_4\Delta r_{n,t} - \underset{(0.01)}{0.06}\ \mathrm{S}_{1973} \\
&+ \underset{(0.03)}{0.06}\ \mathrm{S}_{1979} - \underset{(0.02)}{0.01}\ \mathrm{S}_{1981} + \underset{(0.01)}{0.04}\ \mathrm{S}_{1999} - \underset{(0.02)}{0.04}\ \mathrm{S}_{2008} + \underset{(0.019)}{0.008} \\
\mathsf{R}^2 \;=\;& 0.78\ \ \widehat{\sigma} = 2.82\%\ \ \mathsf{F}_{\mathsf{ar}}(2,115) = 1.04\ \ \mathsf{F}_{\mathsf{arch}}(1,132) = 16.81^{**} \\
\chi^2_{\mathsf{nd}}(2) \;=\;& 20.44^{**}\ \ \mathsf{F}_{\mathsf{het}}(25,108) = 2.15^{**}\ \ \mathsf{F}_{\mathsf{reset}}(2,115) = 9.39^{**}
\end{aligned} \tag{7.9}$$

The estimates are radically altered, the fit is very much worse as judged by $\widehat{\sigma}$, and most mis-specification tests reject at 1% or more.

Figure 7.4 records the graphical statistics, which reveal the main problem is a serious mis-fit around WWI and the early 1920s. Thus, delete the step indicators from (7.9) making sure all regressors are `fixed`, and re-select again using SIS at 0.1% to obtain (7.10).

$$\begin{aligned}
\Delta p_t \;=\;& \underset{(0.06)}{0.30}\ \Delta p_{t-1} + \underset{(0.08)}{0.25}\ \Delta m_t + \underset{(0.03)}{0.09}\ \Delta r_{L,t-1} + \underset{(0.09)}{0.03}\ \Delta m_{t-1} \\
&+ \underset{(0.26)}{0.49}\ e_{t-1} + \underset{(0.01)}{0.02}\ \Delta r_{n,t} + \underset{(0.05)}{0.00}\ \Delta^2(m-p)_{t-2} + \underset{(0.07)}{0.02}\ \Delta g_t \\
&+ \underset{(0.01)}{0.00}\ (D_1+D_3) + \underset{(0.01)}{0.01}\ D_4 - \underset{(0.02)}{0.05}\ D_4\Delta r_{n,t} - \underset{(0.01)}{0.08}\ \mathrm{S}_{1914} \\
&+ \underset{(0.02)}{0.22}\ \mathrm{S}_{1920} - \underset{(0.02)}{0.15}\ \mathrm{S}_{1922} - \underset{(0.01)}{0.07}\ \mathrm{S}_{1973} + \underset{(0.01)}{0.07}\ \mathrm{S}_{1980} + \underset{(0.007)}{0.006} \\
\mathsf{R}^2 \;=\;& 0.89\ \ \widehat{\sigma} = 2.00\%\ \ \mathsf{F}_{\mathsf{ar}}(2,115) = 0.79\ \ \mathsf{F}_{\mathsf{arch}}(1,132) = 6.09^{*} \\
\chi^2_{\mathsf{nd}}(2) \;=\;& 7.89^{*}\ \ \mathsf{F}_{\mathsf{het}}(25,108) = 0.92\ \ \mathsf{F}_{\mathsf{reset}}(2,115) = 0.31
\end{aligned} \tag{7.10}$$

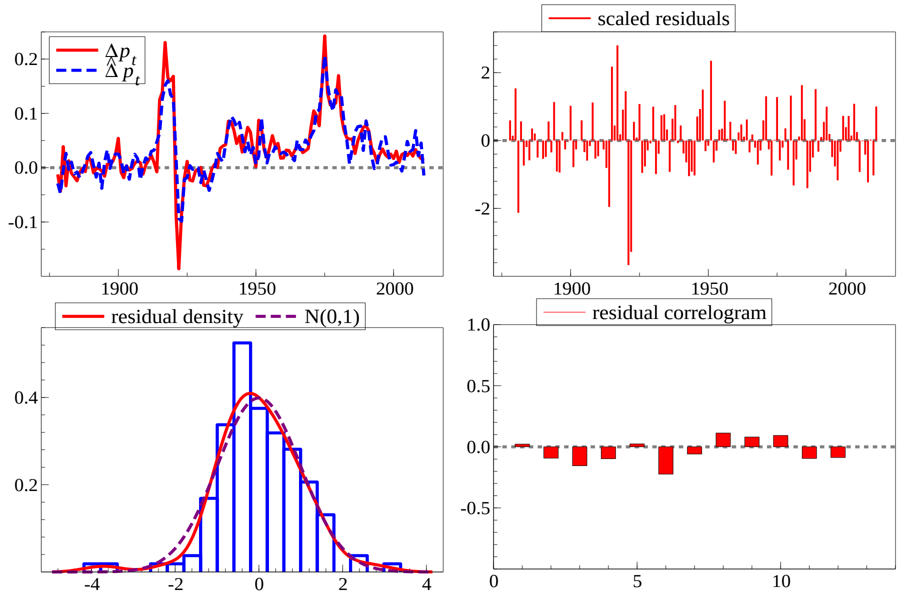


Figure 7.4
Graphical statistics for the money-demand model re-normalized on inflation.

The mis-specification tests are greatly improved, although two remain significant at 5%. However, apart from S_{1973}, all the other selected steps differ from (7.8), and most regressors are now insignificant.

To eliminate the insignificant variables, undo the `fixed` status of all regressors, choose `Automatic model selection` at 1%, setting `Outlier and break detection` to `None`, as we have already found the step indicators. The outcome should be (7.11).

$$\begin{aligned}
\Delta p_t &= \underset{(0.05)}{0.31}\ \Delta p_{t-1} + \underset{(0.04)}{0.28}\ \Delta m_t + \underset{(0.02)}{0.11}\ \Delta r_{L,t-1} - \underset{(0.01)}{0.08}\ S_{1914} \\
&\quad + \underset{(0.02)}{0.23}\ S_{1920} - \underset{(0.02)}{0.16}\ S_{1922} - \underset{(0.01)}{0.06}\ S_{1973} + \underset{(0.01}{0.07}\ S_{1980} \\
(\mathsf{R}^*)^2 &= 0.87\ \ \widehat{\sigma} = 2.06\%\ \ \mathsf{F}_{\mathsf{ar}}(2,124) = 0.89\ \ \mathsf{F}_{\mathsf{arch}}(1,132) = 6.12^* \\
\chi^2_{\mathsf{nd}}(2) &= 7.05^*\ \ \mathsf{F}_{\mathsf{het}}(11,122) = 2.31^*\ \ \mathsf{F}_{\mathsf{reset}}(2,124) = 0.01 \qquad (7.11)
\end{aligned}$$

where $(\mathsf{R}^*)^2$ is the value when a constant is included. Very little remains of the original money demand model, and three diagnostic tests remain significant at 5%.

The next step is to test the invariance of the two representations. First, add the remaining four step indicators S_{1979}, S_{1981}, S_{1999}, S_{2008} from (7.8) to (7.11) and re-estimate. Click on the `Test menu` icon, and tick the box for `Exclusion Restrictions`, `OK`, then highlight the four added step indicators, `OK`, and the test outcome should be $\mathsf{F}(4,122) = 1.36$. Thus, the major unexplained shifts in money did not affect inflation, other than that for 1973 which appears in both models.

You should be able to carry out the converse test for excluding the step indicators S_{1914}, S_{1920}, S_{1922}, S_{1980} in (7.11) from (7.8) to obtain $\mathsf{F}(4,113) = 0.53$. Consequently, Δm_t and Δp_t co-break other than in 1973. However, S_{1973} has a negative effect on Δp_t in (7.11) but was positive and of the same magnitude as the War dummy in (7.8), suggesting it is 'picking up' different effects in the two equations.

7.10 Task 26*: Deriving the impact of error autocorrelation on ESEs

Task 16 Section 4.16 simulated the impact on ESEs of error autocorrelation: here we look at a simple case where it is feasible to analyse the outcome.

Consider the equation:

$$y_t = \beta x_t + u_t \tag{7.12}$$

when both $\{y_t\}$ and $\{x_t\}$ are stationary processes:

$$u_t = \rho u_{t-1} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{7.13}$$

with $|\rho < 1|$ and:

$$x_t = \lambda x_{t-1} + \nu_t \quad \text{where} \quad \nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right] \tag{7.14}$$

also with $|\lambda < 1|$, and $\mathsf{E}\left[\epsilon_t \nu_s\right] = 0 \;\forall t, s$. Then letting $\mathsf{E}[u_t^2] = \sigma_u^2 = \sigma_\epsilon^2 / \left(1 - \rho^2\right)$:

$$\mathsf{E}\left[u_t u_{t-k}\right] = \rho^k \sigma_u^2 \tag{7.15}$$

and as $\mathsf{E}[x_t^2] = \sigma_x^2 = \sigma_\nu^2 / \left(1 - \lambda^2\right)$:

$$\mathsf{E}\left[x_t x_{t-k}\right] = \lambda^k \sigma_x^2 \tag{7.16}$$

When the error autocorrelation in (7.13) is ignored during estimation of β:

$$\widehat{\beta} = \beta + \frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}$$

because of the strong assumption that $\{x_t\}$ is independent of $\{u_t\}$ implied by $\mathsf{E}\left[\epsilon_t \nu_s\right] = 0 \;\forall t, s$, then $\widehat{\beta}$ is unbiased for β conditional on $\{x_t\}$:

$$\mathsf{E}\left[\widehat{\beta}\right] = \beta + \frac{\sum_{t=1}^T x_t \mathsf{E}\left[u_t\right]}{\sum_{t=1}^T x_t^2} = \beta.$$

The variance of $\widehat{\beta}$ is:

$$\mathsf{E}\left[\left(\widehat{\beta} - \beta\right)^2\right] = \mathsf{E}\left[\left(\frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}\right)^2\right] = \mathsf{E}\left[\frac{\sum_{t=1}^T \sum_{s=1}^T \left(x_t x_s u_t u_s\right)}{T^2 \left(T^{-1} \sum_{t=1}^T x_t^2\right)^2}\right] \tag{7.17}$$

Approximate the denominator in (7.17) by $\left(T\sigma_x^2\right)^2$, and use the independence of $\{x_t\}$ and $\{u_t\}$ so that:

$$
\begin{aligned}
\mathsf{E}\left[\left(\widehat{\beta}-\beta\right)^2\right] &\approx \frac{1}{\left(T\sigma_x^2\right)^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathsf{E}\left[x_t x_s\right]\mathsf{E}\left[u_t u_s\right]\right) \\
&= \frac{\sigma_x^2\sigma_u^2}{\left(T\sigma_x^2\right)^2}\left(T+2T\sum_{k=1}^{T}\lambda^k\rho^k\right) \\
&= \frac{\sigma_u^2}{T\sigma_x^2}\left(1+2\lambda\rho\sum_{k=0}^{T-1}\lambda^k\rho^k\right) \\
&\approx \frac{\sigma_u^2}{T\sigma_x^2}\left(1+\frac{2\lambda\rho}{(1-\lambda\rho)}\right) \\
&= \frac{\sigma_u^2\left(1+\lambda\rho\right)}{T\left(1-\lambda\rho\right)\sigma_x^2}
\end{aligned} \tag{7.18}
$$

The estimated standard error (ESE) of $\widehat{\beta}$ is is the square root of its conventionally calculated variance given by:

$$
\mathsf{Var}\left[\widehat{\beta}\right] = \frac{\sigma_u^2}{\sum_{t=1}^{T}x_t^2} \approx \frac{\sigma_u^2}{T\sigma_x^2} \tag{7.19}
$$

which under-estimates $\mathsf{E}[(\widehat{\beta}-\beta)^2]$ in (7.18) by $\left(1+\lambda\rho\right)/\left(1-\lambda\rho\right)$. For example, if $\lambda=\rho=0.9$, that is more than 9-fold, so the t-ratios are inflated 3-fold! However, the bias in (7.19) for (7.18) vanishes if either λ or ρ is zero.

All three results match the Monte Carlo in Task 16 Section 4.16, where the ratio of the simulation MCSD to the average ESE, was three, and the ratio was unity if either λ or ρ was zero.

7.11 Task 27: Using lags to eliminate residual autocorrelation

Set the `Category` to `Models for time-series data`, and the `Model class` to `Single-equation Dynamic Modelling using PcGive`.
Click on `Formulate`, set `lags` to 0 and select m, then `OK`, `OK`, `OK`, to estimate m_t regressed on a constant. Click on the `Test Menu` icon (second last on the right on the Icon line, or Alt+**T**), and tick the `Graphic Analysis...` box, then `OK` for the `Graphic Analysis` dialog. Untick the two boxes for `Actual and fitted values` and `Residuals (scaled)` and tick the box for `Residual autocorrelations (ACF)`. The residual correlogram is everywhere high and positive, declining slowly (the actual successive correlations are very close to unity even for 12 years previously), as seen in Figure 7.5 panel *a*. Copy the correlogram to a new data plot. Repeat, but with `lags` set to 1, then 2, and finally 4. When m_t is regressed on m_{t-1} and a constant as in panel *b*, the residual autocorrelations soon change sign and become small. Extending the model to m_t being regressed on $m_{t-1},\ldots,m_{t-4}$ and a constant, the residual autocorrelations are all small as seen in panel *d*, and are not significantly different from zero, even though m_t trends strongly and is obviously non-stationary.

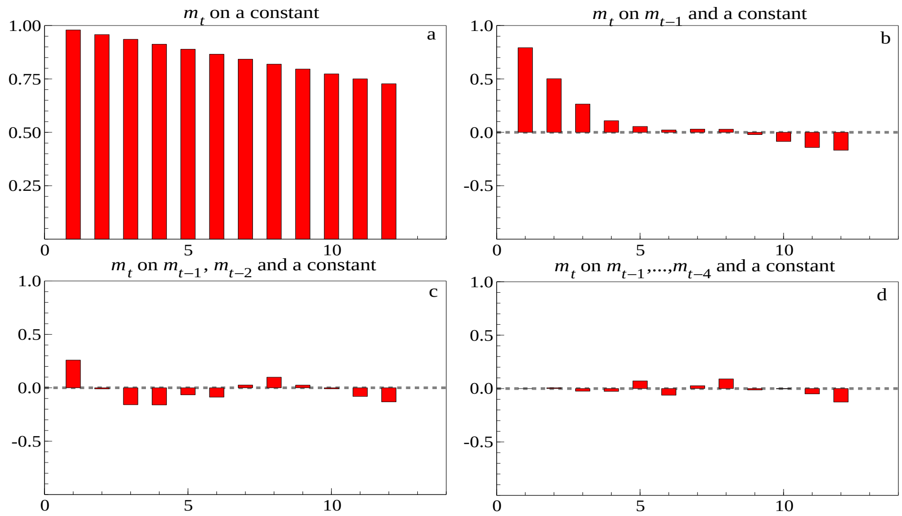


Figure 7.5
Nominal money residual correlograms

It is intuitively clear why additional lags can reduce residual autocorrelations. For example, substituting (4.53) into (4.54) delivers:

$$y_t = \beta x_t + \rho y_{t-1} - \beta\rho x_{t-1} + \epsilon_t \tag{7.20}$$

so adding both lagged variables induces the independent error ϵ_t.

7.12 Task 28*: Sequential factorization

However, there is a more fundamental result underlying outcomes like Figure 7.5, called sequential factorization. This generalizes a well-known result about the probability $\mathsf{P}(A, B)$ of two events, A and B, say, which can be written as:

$$\mathsf{P}(A, B) = \mathsf{P}(A \mid B)\mathsf{P}(B) \tag{7.21}$$

where $\mathsf{P}(A|B)$ is the probability of A happening given that B has occurred, times the probability of B happening. In (7.21), $\mathsf{P}(A|B)$ is the conditional probability, and B is the marginal probability. So in the UK, for example, the probability of being rich and young is the probability of being rich when young, times the probability of being young, and the names derive from tables of distributions of income by age, where the margins of the table recorded the distributions of income and of age, and the rows in the table showed the incomes at any given age.

The result in (7.21) can be extended to a series of probabilities:

$$\mathsf{P}(A, B, C) = \mathsf{P}(A \mid B, C)\mathsf{P}(B|C)\mathsf{P}(C) \tag{7.22}$$

where C might denote being female.

When $\{y_t\}$ is a stationary process over a long history, $t = \ldots - 2, -1, 0, 1, 2, \ldots$, then $(y_1, y_2, \ldots, y_{T-1}, y_T)$ can be interpreted as a single observation on that process, with a common mean, $\mathsf{E}[y_t] = \mu$, variance $\mathsf{Var}[y_t] = \sigma_y^2$, and autcorrelations

$\mathsf{E}[y_t y_{t-k}] = \rho_k$, with $|\rho_k| < 1$. Consider 'explaining' y_t by its conditional expectation given its history, $(y_{t-1}, y_{t-2}, y_{t-3}, \ldots)$:

$$y_t = \mathsf{E}[y_t \mid y_{t-1}, y_{t-2}, y_{t-3}, \ldots] + \epsilon_t \tag{7.23}$$

Taking expectations on both sides conditional on $(y_{t-1}, y_{t-2}, \ldots)$:

$$\mathsf{E}[y_t \mid y_{t-1}, y_{t-2}, \ldots] = \mathsf{E}[y_t \mid y_{t-1}, y_{t-2}, \ldots] + \mathsf{E}[\epsilon_t \mid y_{t-1}, y_{t-2}, \ldots] \tag{7.24}$$

so that $\mathsf{E}[\epsilon_t | y_{t-1}, y_{t-2}, \ldots] = 0$. Thus, $\{\epsilon_t\}$ cannot be explained by the past any better than its mean of zero, and as $\epsilon_t = y_t - \mathsf{E}[y_t \mid y_{t-1}, y_{t-2}, \ldots]$ from (7.23), it is not autocorrelated: $\mathsf{E}[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots] = 0$, so we have created a 'random' error.[1]

This process can be applied at every t, so we can write the joint distribution, $\mathsf{D}_y(\cdot)$ of $(y_1, y_2, \ldots, y_{T-1}, y_T)$ in reverse time order as:

$$\begin{aligned}
\mathsf{D}_y\left(y_T, y_{T-1}, \ldots, y_2, y_1, y_0\right) &= \mathsf{D}_y\left(y_T \mid y_{T-1}, \ldots, y_2, y_1, y_0\right) \times \\
&\quad\ \mathsf{D}_y\left(y_{T-1}, \ldots, y_2, y_1, y_0\right) \\
&= \mathsf{D}_y\left(y_T \mid y_{T-1}, \ldots, y_2, y_1, y_0\right) \times \\
&\quad\ \mathsf{D}_y\left(y_{T-1} \mid y_{T-2}, \ldots, y_2, y_1, y_0\right) \times \\
&\quad\ \mathsf{D}_y\left(y_{T-2}, \ldots, y_2, y_1, y_0\right) \\
&= \qquad \vdots \\
&= \prod_{t=1}^{T} \mathsf{D}_y\left(y_t \mid y_{t-1}, \ldots, y_2, y_1, y_0\right)
\end{aligned} \tag{7.25}$$

where $\prod_{t=1}^{T}$ denotes the product of the conditional distributions. Now we can apply (7.23) to each of these conditional distributions $\mathsf{D}_y(y_t | y_{t-1}, \ldots, y_2, y_1, y_0)$, which have means $\mathsf{E}[y_t | y_{t-1}, y_{t-2}, \ldots]$, the deviations from which are the not autocorrelated errors $\{\epsilon_t\}$. Such a result holds irrespective of the extent of the original autocorrelations $\mathsf{E}[y_t y_{t-k}] = \rho_k$ between the y_t.

7.13 Chapter 7 exercises

1. Repeat Task 27 Section 7.11 for $(m-p)_t$ over a sample ending in 2011 for regressions on a constant and 0, 1, 2, and 4 lags, collecting the four residual correlogram plots. Compare the outcomes with those shown in Figure 7.5 for m_t.
2. Next, repeat *i*) with all regressors marked as fixed over the same sample ending in 2011, but select by SIS at 0.1%. Paste each residual correlogram onto that in *i*) to compare the impact of SIS on residual autocorrelation (see Figure 7.6).
3. What, if anything, can be deduced about the impact of location shifts on residual autocorrelation ?
4. Calculate the correlograms for 'Dmp' ($\Delta(m-p)_t$) and 'DDmp' (namely $\Delta^2(m-p)_t$) and compare them to those from the regressions of $(m-p)_t$ on $(m-p)_{t-1}$ and a constant, and $(m-p)_t$ on $(m-p)_{t-1}$, $(m-p)_{t-2}$ and a constant respectively. Explain any similarities and any large differences.
5. Finally, calculate the residual correlograms from separate regressions for $\Delta(m-p)_t$ on a constant marked as fixed with IIS and SIS and discuss the results.

[1] Technically, $\{\epsilon_t\}$ is called a martingale difference sequence, and such processes are important in finance.

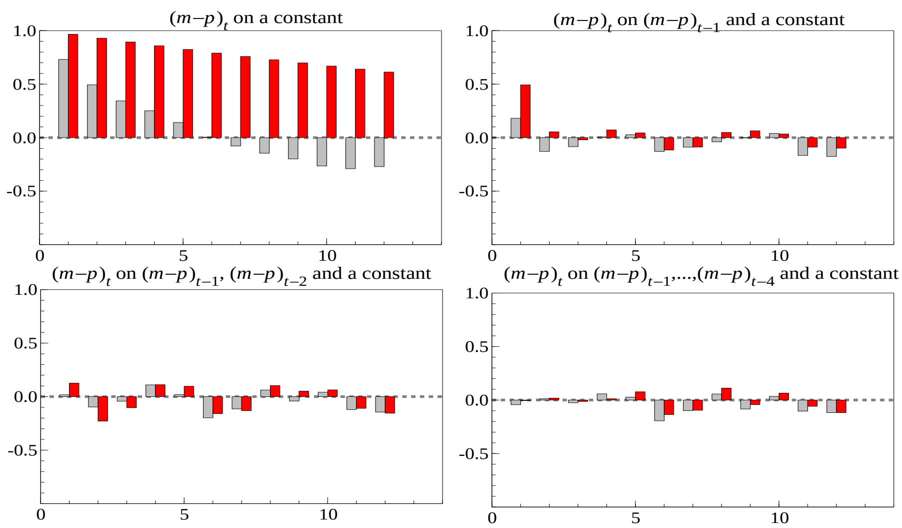


Figure 7.6
Real money residual correlograms without SIS (dark) and with (light)

Chapter 8
Modelling UK prices

Chapter 8 guide posts

1. There are many potential theory-relevant domestic and foreign determinants of inflation: Section 8.1.
2. A previous empirical analysis starting from a GUM that included all theory-suggested variables and their dynamics found most such variables were relevant yet the model still had many large shifts: Section 8.2.
3. To update and extend that equation to 2011 requires modeling all the excess demand influences on inflation, and investigating their importance: Section 8.3.
4. Location shifts are reasonably captured by impulse and step indicators: Section 8.4.
5. The price inflation model has two equilibrium corrections, one from the excess demand for goods and services, and the other from the price markup over home and imported costs and despite the emphasis in some theories of the role of unemployment in the price inflation process little evidence of its importance is found: Section 8.5.
6. Also, expectations of future inflation are not found to matter greatly: Section 8.6.
7. Overall, viable empirical models of price inflation can be developed despite its turbulent history.

8.1 Price inflation determinants

We have noted many major differences between life in 1860 and that over the ensuing 150 years, and the final one concerns inflation. During a working life from 1860 to 1900, as Figure 2.1 showed, the price level would have risen by less than 10%, and most of that rise occurred in 1900 itself (probably from demand pressures during the Boer War). During the equivalent 1960–2000, the price level rose more than 1300%. To complete the explanation of the nominal level in the UK economy, we need to model prices.

We addressed the theory that inflation is primarily a monetary phenomenon in Chapter 7 by modelling UK money demand using the broad money measure M4, and showing that ‘inverting’ the status of money and prices to explain inflation led to a poor model. In a modern economy, price inflation seems to be the resultant of all the excess demands in the economy, from money, cost push, especially from wages, demand pull, devaluation, and profligate governments. No single factor seems an adequate explanation, so as with modelling wages, we will include them all. This necessitates measures for:

(a) excess demands for goods and services;
(b) excess money holding;
(c) excess demands for factors of production;
(d) major changes to exchange rates and the prices of imports;
(e) excess government demands (unfunded deficits);
(f) special factors such as wars, commodity prices bubbles, and wage or price controls.

We will measure these respectively by:
(a) the deviation of output from capacity (the ‘gap’);
(b) excess money supply (modelled in Chapter 7);
(c) nominal wage inflation, unemployment, and the markup of prices over costs;
(d) purchasing power parity deviations and imported inflation;
(e) national debt, and long- and short-run interest rates;
(f) indicator variables, commodity prices, and lagged rates of change.

8.2 An earlier price inflation model

The price inflation model in (8.1) is taken from Hendry (2001), and was selected before IIS was available: the indicators used are discussed in section 8.2.1.

$$\begin{aligned}\Delta p_t &= \underset{(0.03)}{0.18}g^d_{t-1} - \underset{(0.02)}{0.19}\pi^*_{t-1} - \underset{(0.09)}{0.83}(R_S - R_L + .0065)_{t-1} \\ &\quad + \underset{(0.11)}{0.62}\Delta R_{S,t-1} + \underset{(0.03)}{0.19}\Delta m_{t-1} + \underset{(0.03)}{0.27}\Delta p_{w,t} \\ &\quad + \underset{(0.03)}{0.27}\Delta p_{t-1} + \underset{(0.01)}{0.04}\Delta p_{o,t-1} + \underset{(0.002)}{0.04}\, I_{d,t} \qquad (8.1) \\ \mathsf{R}^2 &= 0.975 \quad \widehat{\sigma} = 1.4\% \quad T = 1875 - 1991\end{aligned}$$

In (8.1), g^d_t denotes the excess demand for goods and services calculated as the deviation of g_t from the estimated production function approximating Figure 1.9, with changing rates of technical progress. π^* denotes the markup of prices over ulc^*, nominal unit labour costs adjusted for hours, computed as:

$$\pi^*_t = \left(p_t - 0.675ulc^*_t - 0.25p_{£,t} - 0.075p_{o,t}\right) + 0.11I_{2,t} + 0.25. \qquad (8.2)$$

where $p_{£,t} = p_{w,t} - e_t$ denotes world prices in sterling, $\Delta p_{w,t}$, $\Delta p_{o,t}$ are world price and oil price inflation, and $I_{2,t}$ is an indicator equal to unity after WWII. Finally, $I_{d,t}$ is a composite indicator which we will now discuss.

8.2.1 Impulse indicators for price inflation

Prior to the invention of IIS, impulse indicators were based on institutional knowledge of turbulent periods and outliers (defined as greater than $3\widehat{\sigma}$ here), which are observations that remain discrepant even after all the explanatory variables are included. Indicator variables were needed for huge, large, and medium outliers, roughly 12%, 6% and 4% respectively, making 22 in total as noted in the Appendix, then were combined into an overall index, I_d, with weights of 3, 1.5, and unity:

$$I_{d,t} = 3I_{b,t} + 1.5I_{L,t} + I_{m,t} \tag{8.3}$$

Such a plethora of indicators reveals that the largest shifts in UK price inflation are not explained by any variables suggested from the wide range of economic theories discussed above—and the model still needed $I_{2,t}$. Although indicators will reflect large measurement errors as well as shifts, only 1880–1881 does not correspond to an important historical event, so may reflect measurement errors.

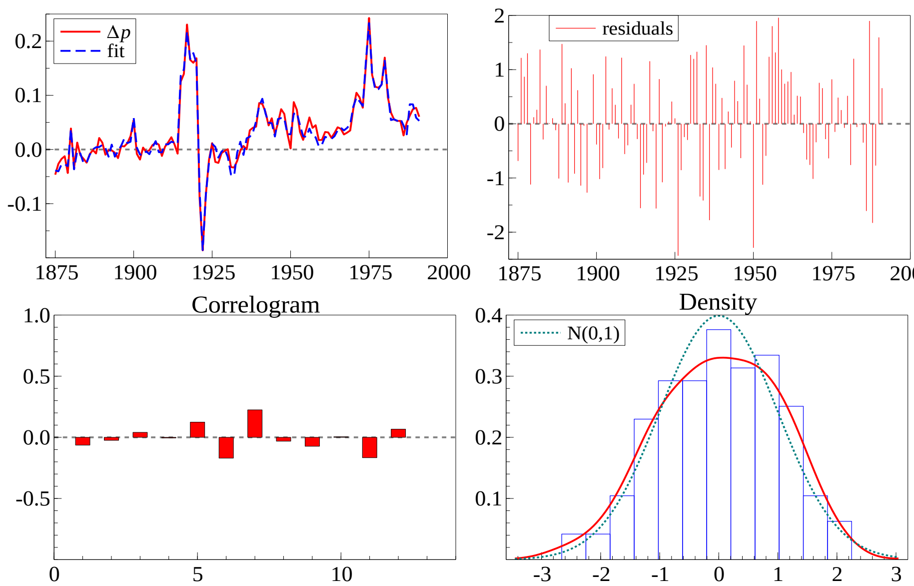


Figure 8.1
Price inflation model (8.1) graphics

Figure 8.1 records the usual graphical statistics. The close fit, with homoskedastic, serially uncorrelated, near normal residuals, is due in no small way to the many indicators used. Dropping $I_{d,t}$ from (8.1) increases the residual standard deviation markedly to $\widehat{\sigma} = 0.029$.

8.3 Updating and extending the price inflation model

Given the 20 new observations, and the advances in econometric technology since (8.1) was modelled, we will now develop an extended version. That requires selecting 'auxiliary' models for the price inflation determinants (a)–(f) in section 8.1 as follows.

8.3.1 (a): deviation of output from capacity

The 'production function' for $(g-l)_t$ discussed in section 1.7 was modelled as follows. The GUM was specified as a generalization of (1.6) with four lags of $(g-l)_t$ and $(k-wpop)_t$, where the log-ratio of capital to the working population was used to avoid sudden shifts due to changes in unemployment. Those regressors, constant and trend were fixed, and SIS implemented for 1862–1944 at 0.1%. Then the lags were unrestricted, with selection at 1%, which ensured the 'static' production function was retained. The same procedure was then used for the period 1945–2004.

The resulting equations were similar to those in section 3.1 of Hendry (2001).

$$
\begin{aligned}
(g-l)_t &= \underset{(0.068)}{0.705}\,(g-l)_{t-1} + \underset{(0.055)}{0.083}\,(k-wpop)_t + \underset{(0.27)}{0.651} \\
&\quad + \underset{(0.0009)}{0.002}\,t - \underset{(0.015)}{0.116}\,\mathrm{S}_{1918} - \underset{(0.015)}{0.092}\,\mathrm{S}_{1920} \\
\mathsf{R}^2 &= 0.992 \quad \widehat{\sigma} = 1.89\% \quad \mathsf{F}_{\mathsf{ar}}(2,75) = 1.44 \quad \mathsf{F}_{\mathsf{arch}}(1,81) = 0.79 \\
\chi^2_{\mathsf{nd}}(2) &= 2.93 \quad \mathsf{F}_{\mathsf{het}}(8,74) = 0.52 \quad \mathsf{F}_{\mathsf{reset}}(2,75) = 3.18^{*}
\end{aligned}
\tag{8.4}
$$

leading to the long-run solution (without step indicators), where $\mathrm{t}_{ur} = -4.34^{*}$:

$$
e_{gl1,t} = (g-l)_t - 2.21 - 0.0068t - 0.28\,(k-wpop)_t \tag{8.5}
$$

For 1945–2004, the static production function with location shifts yielded:

$$
\begin{aligned}
(g-l)_t &= \underset{(0.31)}{1.69} + \underset{(0.049)}{0.39}\,(k-wpop)_t + \underset{(0.0013)}{0.015}\,t + \underset{(0.01)}{0.035}\,\mathrm{S}_{1946} \\
&\quad + \underset{(0.008)}{0.044}\,\mathrm{S}_{1973} + \underset{(0.008)}{0.028}\,\mathrm{S}_{1978} + \underset{(0.007)}{0.042}\,\mathrm{S}_{1989} + \underset{(0.007)}{0.016}\,\mathrm{S}_{1995} \\
\mathsf{R}^2 &= 0.999 \quad \widehat{\sigma} = 1.23\% \quad \mathsf{F}_{\mathsf{ar}}(2,50) = 5.01^{*} \quad \mathsf{F}_{\mathsf{arch}}(1,58) = 0.32 \\
\chi^2_{\mathsf{nd}}(2) &= 0.42 \quad \mathsf{F}_{\mathsf{het}}(9,50) = 1.24 \quad \mathsf{F}_{\mathsf{reset}}(2,50) = 1.06
\end{aligned}
\tag{8.6}
$$

leading to (including but not reporting the step indicators):

$$
e_{gl2,t} = (g-l)_t - 1.69 - 0.015t - 0.39\,(k-wpop)_t \tag{8.7}
$$

This was then extended to 2011. Capacity (denoted cap_t) was calculated by combining the right-hand sides of (8.5) and (8.7), so the measure of excess demand for goods and services, g^d_t, is $e_{gl1,t}$, $e_{gl2,t}$ combined.

Figure 8.2 panel a, shows the plot of cap_t with $(g-l)_t$ and panel b the resulting excess demand (or supply), g^d_t. The depressions at the start of period, the inter-war epoch and the 'Great Recession' from 2008 are very visible, as are the peaks of the Boer War, WWI and WWII, together with the so-called 'stop-go' economic policies of the post-war till the mid 1970s.

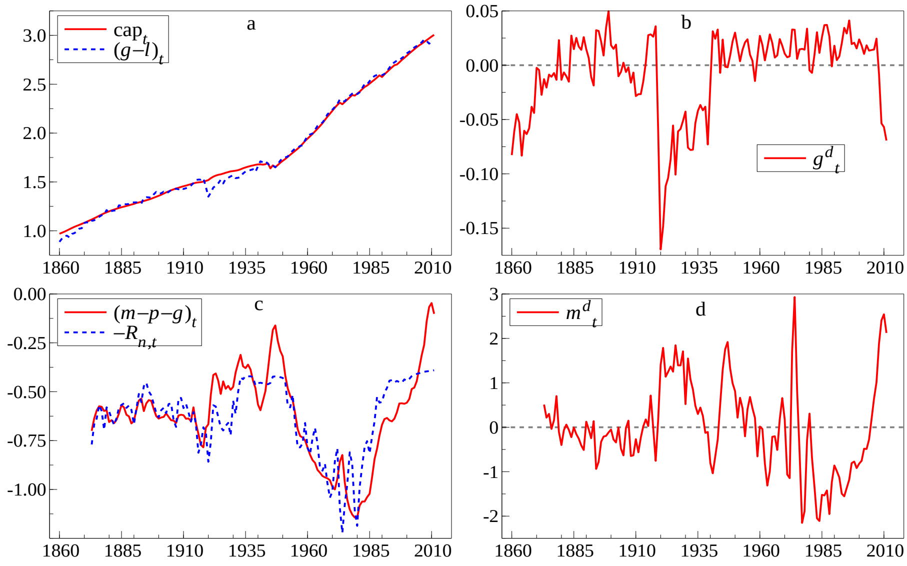


Figure 8.2
Capacity, excess demand for goods and services, velocity and interest rates, and excess demand for money

8.3.2 (b) excess demand for money

The excess demand for money was computed as $m^d = \widetilde{u}$ from (7.6). Figure 8.2 panel c shows the graph of inverse velocity (i.e., $(m - p - g)_t$) and the negative of $R_{n,t}$ (matched as $-7.4R_{n,t}$), and panel d records m^d_t. The increase in m^d_t starting from the trough in the mid 1980s has been very large. As discussed in Chapter 7, both Ericsson, Hendry, and Prestwich (1998) and later Escribano (2004) showed that the earlier dynamic model of UK money demand in Hendry and Ericsson (1991) remained constant for their extended samples, updated to 2011 above, although the latest period required additional step indicators.

8.3.3 (c) unit labour costs and the markup

Unit labour costs $c_t = (w + l - g)_t$ were introduced in Section 2.6.2, and formed a key component of the markup π^*_t adjusted for hours in (8.2), which was highly significant in (8.1). Here we use essentially the same formulation:

$$\pi_t = p_t - 0.675\,(c_t + 0.005t) - 0.25p_{£,t} - 0.075p_{o,t} + 0.05I_{2,t} - 1.55. \qquad (8.8)$$

where $p_{£,t} = p_{w,t} - e_t$ as before, the implicit adjustment for changes in hours is 0.5% p.a. and the remaining coefficients are as in (8.2): the constant reflects revisions and changes in base years, and $I_{2,t}$ has a smaller effect.

Figure 8.3 panel a, plots π_t. The sharp drop in the markup since 2000 is consistent with competition from China, and the lower rate of inflation the UK has experienced this Century compared to last, closer to 19th Century experiences.

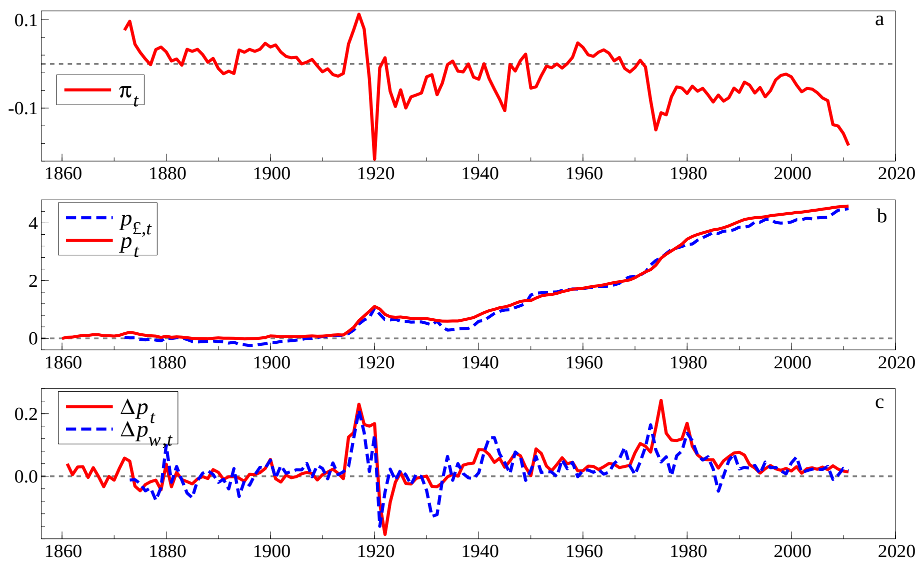


Figure 8.3
Markup, UK prices, world prices in Sterling, and inflation rates, Δp_t and $\Delta p_{w,t}$

8.3.4 (d) world price inflation and PPP deviations

Figure 2.8 presented graphs for the £ sterling exchange rate index, e_t, the logs of the UK price level, p_t, and world prices, $p_{w,t}$ (middle panel), and the purchasing power parity given by $ppp_t = e_t + p_t - p_{w,t}$.

Here, Figure 8.3 records the comparative levels of UK and world prices in Sterling (panel b), and the inflation rates, Δp_t and $\Delta p_{w,t}$ (panel c). Expressed in a common currency, p_t and $p_{£,t}$ behave quite similarly, although in their own currencies, UK and world price inflation exhibit many substantial departures. However, $\Delta p_{w,t}$ helps explain Δp_t below.

8.3.5 (e) interest rates and National Debt

Figure 2.9 showed the UK's long-term and short-term interest rates, R_L and R_S, and their spread, $R_L - R_S$. However, the role of National Debt is surprisingly hard to ascertain econometrically. Figure 8.4 records four transformations of potential relevance: $(N/PG)_t$ (panel a), Δn_t (b), $\Delta(n-p)_t$ (c) and $\Delta(n-p-g)_t$ (d).

The ratio of National Debt to nominal National Income $(N/PG)_t$ in panel a is shown from 1820, and is back to levels around the middle of the 19th Century when Debt accumulated during the Napoleonic Wars was being repaid. Nevertheless, $(N/PG)_t$ remains well below its peak values, which were eliminated quite quickly during the 1950s as nominal GDP grew relatively rapidly.[1] Indeed, $(N/PG)_t$ is below its historical average, although at much higher levels, the UK was able to finance the Industrial Revolution (see Allen, 2009), and build a large empire. The growth of nominal National Debt Δn_t is primarily associated with War Finance:

[1] For a discussion of Net (as against Gross here) National Debt up to the late 1970s, see Hendry, 1980.

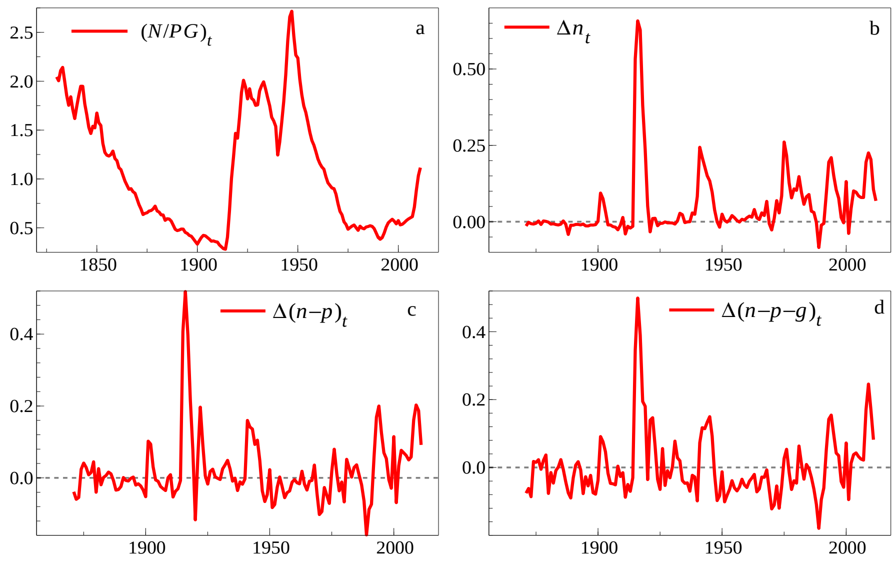


Figure 8.4
UK National Debt

impulses at the Boer War, WWI, and WWII are manifest in the period before Governments attempted to stabilize the economy. Thereafter, the next three fast, and relatively similar, increases are during the 1974–1978 oil Crisis, then the 1990–1995 recession under John Major, and the Great Recession from 2007. Given the dramatic fall in GDP, the last increase is actually surprisingly small. In constant prices, the last two rises are similarly fast, and comparable to WWII. Relative to GDP, the most recent rise is the fastest, partly because GDP fell considerably, but no role was found for National Debt in explaining UK inflation.

8.3.6 (f) commodity prices and special events

Figure 8.5 panels a and b show the log level and changes in an index of nominal raw material prices (including fuel).

Nominal price changes have been large, with WWI, the post-war crash, the Great Depression, WWII, the two Oil Crises, the rise of China during the first decade of the 21st Century, and the Great Recession all too visible, albeit not the only times of large shifts. Figure 8.5 panels c and d show the levels of nominal and real commodity prices measured in £: the former has risen much more than $ prices because of the relative fall in the value of the £, whereas the latter has not really trended since the end of the 19th Century, with a marked recent recovery, probably due to China's large demands to sustain its rapid growth.

A price index for commodities raises an interesting econometric issue, namely the meaning of a constant relationship. By construction:

$$p_{o,t} = \sum_{i=1}^{n} w_{i,t} p_{i,t}$$

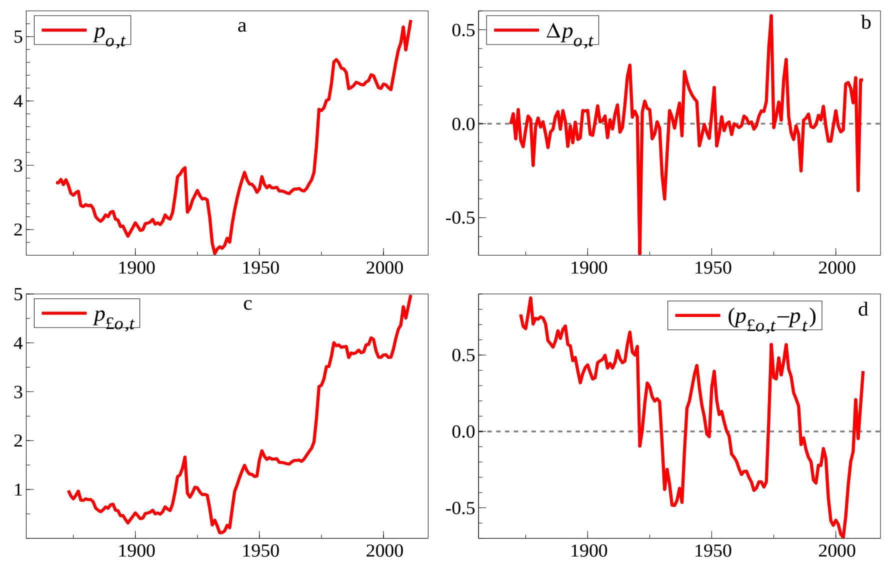


Figure 8.5
Commodity prices

where the $\{p_{i,t}\}$ are the prices of individual commodities, including many different metals, fibres, foods, and fuels. The $\{w_{i,t}\}$ are their weights in the index, often based on value shares, sometimes on trade weights. In 1860, at the start of our period, various commodities were very important, such as coal and whale oil in fuels, whereas others like petroleum and natural gas, were not yet so important. By 2011, the weights of most will have changed substantively. The impact of a change in the price of any commodity on overall inflation will have altered with its importance in production and consumption, so it seems unlikely that any individual $p_{i,t}$ would have a constant relation to p_t over 150 years. However, that does not by itself preclude a constant relationship between $p_{o,t}$ and p_t, where all the change is 'hidden' in the shifting weights $\{w_{i,t}\}$. More general settings are discussed in Hendry (1996).

8.3.7 Measurement errors

Economic data seek to accurately measure the corresponding variables that arise in theoretical reasoning. Thus, Δp_t here seeks to measure 'UK inflation'. In practice, there are many concepts of inflation and measures thereof, most being implicit deflators of aggregates like GDP, and subject to the analysis in the previous section. Over very long historical periods, measurement errors relative to the underlying 'reality' must be common, especially as the data were rarely recorded initially, and have been reconstructed since from archival information. An example here is that unemployment statistics only use Trades Unions membership information. Incorrect measurements can bias parameter estimates, induce autocorrelated errors in dynamic models, and can lead to non-constancy in estimated equations.

Earlier chapters showed that the levels of most of the time series are I(1) with location shifts, so have very large and non-constant data second moments. Thus, even if the data were incorrectly measured by I(1) measurement errors, changes in both data and measurement errors will be I(0), but the data will remain perturbed by breaks, so the 'signal-noise' ratio will remain high. Of course, changes to the measurement system and its accuracy induce a further non-stationarity for the econometrician to handle. Revisions add further to the practical problems that must be addressed (Hendry, 1995a, ch.14, discusses the impacts of revisions on post-war quarterly inflation time series). Nevertheless, as noted above, analyses of other long historical time series have yielded insights into money demand, and established constant-parameter relationships. Thus, the available empirical evidence merits modelling even if there are *caveats* about measurement accuracy.

8.4 The updated and extended inflation model

Over the same sample period 1875–1991 as (8.1), that equation could be closely replicated on the revised and extended data as follows, using π_{t-1} rather than π^*_{t-1} as noted.

$$\begin{aligned}
\Delta p_t &= \underset{(0.04)}{0.19} g^d_{t-1} - \underset{(0.04)}{0.24}\pi_{t-1} - \underset{(0.09)}{0.74}(R_S - R_L + .0065)_{t-1} + \underset{(0.03)}{0.19}\Delta m_{t-1} \\
&\quad + \underset{(0.12)}{0.59}\Delta R_{S,t-1} + \underset{(0.03)}{0.28}\Delta p_{w,t} + \underset{(0.03)}{0.27}\Delta p_{t-1} + \underset{(0.01)}{0.03}\Delta p_{o,t-1} + \underset{(0.002)}{0.04} I_{d,t} \\
(\mathsf{R}^*)^2 &= 0.962 \;\; \widehat{\sigma} = 1.30\% \;\; SC - 8.43 \;\; \mathsf{F}_{\mathsf{ar}}(2, 106) = 0.39 \\
V &= 0.14 \;\; J = 1.57 \;\; \mathsf{F}_{\mathsf{arch}}(1, 115) = 0.51 \\
\chi^2(2) &= 0.14 \;\; \mathsf{F}_{\mathsf{reset}}(2, 106) = 2.0 \;\; \mathsf{F}_{\mathsf{het}}(18, 98) = 2.3^{**} \qquad (8.9)
\end{aligned}$$

The test for non-linearity in Castle and Hendry (2010) yielded $\mathsf{F}_{\mathsf{nl}}(27, 81) = 1.72^*$. Despite the similarities in the estimated coefficients, the heteroskedasticity and non-linearity tests reveal that the data revisions and different measures of the constructed excess demand and markup have somewhat altered the relationship.

Thus, retaining all of its variables, (8.9) was selected over the sample 1875–1999 using SIS at 0.5% (a technique that was not available when the earlier model was selected), leading to the retention of 7 step indicators, for 1936 (−1.7%), 1949 (3.8%), 1950 (−7.0%), 1951 (3.8%), 1973 (−3.3%), 1975 (3.0%), and 1993 (2.0%), only the last lying outside the earlier sample: these are denoted $\{S_{i,t}\}$ below. Finally, the equation was estimated to 1999, so the data for the 21st Century could be used as a parameter constancy 'forecast test' over 2000–2011.

$$\begin{aligned}
\Delta p_t &= \underset{(0.03)}{0.12}\; g^d_{t-1} - \underset{(0.04)}{0.13}\pi_{t-1} - \underset{(0.10)}{0.54}(R_S - R_L + .0065)_{t-1} \\
&\quad + \underset{(0.03)}{0.14}\Delta m_{t-1} + \underset{(0.11)}{0.47}\Delta R_{S,t-1} + \underset{(0.03)}{0.25}\Delta p_{w,t} \\
&\quad + \underset{(0.03)}{0.29}\Delta p_{t-1} + \underset{(0.01)}{0.03}\Delta p_{o,t-1} + \underset{(0.002)}{0.04}\, I_{d,t} + \{S_{i,t}\} \qquad (8.10) \\
(\mathsf{R}^*)^2 &= 0.968 \;\; \widehat{\sigma} = 1.1\% \;\; SC - 8.49 \;\; \mathsf{F}_{\mathsf{ar}}(2, 107) = 0.09 \\
\chi^2(2) &= 3.6 \;\; \mathsf{F}_{\mathsf{arch}}(1, 123) = 0.85 \;\; \mathsf{F}_{\mathsf{reset}}(2, 107) = 4.0^* \\
\mathsf{F}_{\mathsf{het}}(23, 99) &= 1.42 \;\; \mathsf{F}_{\mathsf{nl}}(27, 82) = 1.38 \;\; \mathsf{F}_{\mathsf{chow}}(12, 109) = 0.97
\end{aligned}$$

The improved fit is mainly due to detecting the earlier step shifts using SIS (although the three for 1949, 1950 and 1951 essentially combine to impulses), and the earlier heteroskedasticity and non-linear test rejections have been removed (a constant is insignificant if added).

The outcome in (8.10) is consistent with a more benign inflation environment since the previous study ended, where 22 large or very large impulses had needed indicators. In late 1992, the UK left the European Exchange Rate Mechanism (ERM) on what is often called Black Wednesday, but transpired to be the start of one of the longest and fastest periods of economic growth till 2008 when the world-wide Financial Crisis and Great Recession hit. The step indicator up to 1993 of 2.0% is consistent with a much lower inflation rate thereafter: the Bank of England was granted independence in 1998 to follow an inflation target of around 2.5%, and there was considerable downward pressure on prices from the rapid growth of China's economy and its competitive exports. Even with these important changes, the estimated model in (8.1) has not shifted substantively over the 20 years since it was first developed. Although quite a few of the final 'forecasts' are over-estimates, so inflation has been somewhat lower than might otherwise have been anticipated, the Chow test value is less than unity, so the fit is better to this last period than on average over the sample with indicators. All but the last forecast lie well within the 95% intervals shown as error bands.

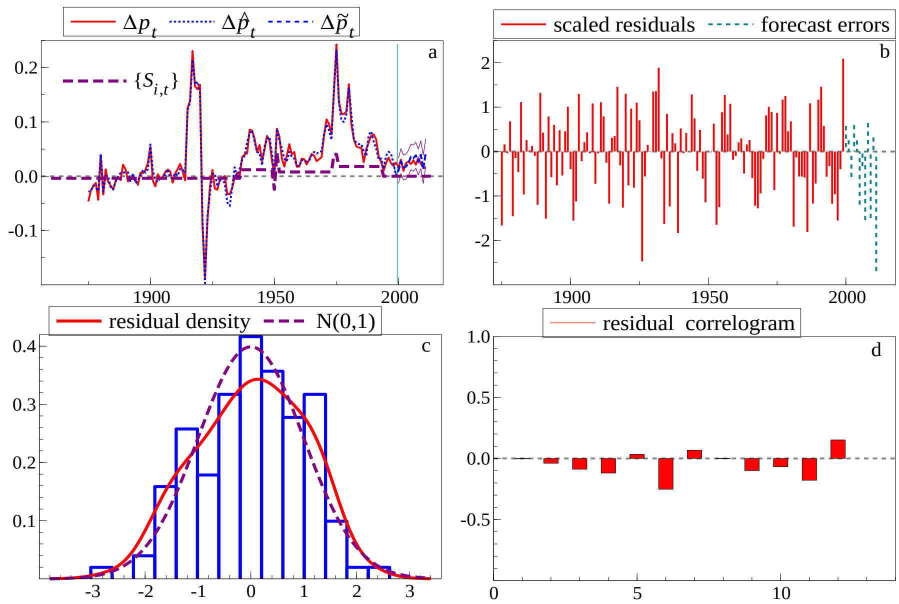


Figure 8.6
Price inflation model (8.10) graphical statistics

Figure 8.6 records the graphical statistics, which are consistent with the more formal tests above.

8.5 Interpreting the inflation model

To put the impact of the step indicators in context, panel a shows their combined effect in $\{S_{i,t}\}$ as a dashed line: as can be seen, despite finding seven significant at 0.5%, the step indicators account for little of the explanation, mainly a slightly higher average inflation rate over the post-war period till 1993. In contrast, the 22 impulse indicators in $I_{d,t}$ represented very large jumps and drops, from 12% for the largest to 4% for the smallest, all at what would otherwise be outlier residuals. Their role is not only to adjust for aspects of Δp_t that are not explained by the regressors, but also correct for shifts in the regressors that do not match those of the dependent variable, although here the former role seems dominant.

Most theories of inflation transpire to be part of the overall explanation in (8.1) and (8.10). The significant proximate determinants include the excess demand for goods and services (g^d_{t-1}, representing (a) above), nominal money growth (Δm_{t-1}, b), the markup of prices over costs (π_{t-1}, c), world price inflation ($\Delta p_{w,t}$, d), the long-short interest rate spread and short-term interest rate changes ($(R_S - R_L + 0.0065)_{t-1}$ and $\Delta R_{S,t-1}$, e), and commodity prices and special events ($\Delta p_{o,t-1}$, plus many step and impulse indicators, f). Indicators for many special events remain crucial despite all these theoretically-suggested variables. Several important episodes are not explained, especially the high inflation during the First World War, followed by the collapse in 1920–22, the rapid inflation in 1940 then low inflation for the rest of World War II (almost certainly reflecting the success of rationing and price controls), step shifts after that war, and the high inflation during the 1970's oil crises 'stagflation', with only the departure from the ERM in the 30 years since. Finally, inflation is not highly inertial, as the lagged value has a coefficient of only 0.29, and although there are seven step indicators implicit in (8.10), there were none in (8.1) where lagged inflation had a similar coefficient.

The price inflation model in (8.10) has two equilibrium corrections, one from the excess demand for goods and services, and the other from the price markup over home and imported costs. The first will affect inflation so long as $g^d \neq 0$ or rounding and setting $u_r = \log(0.05) \approx -3$, until $g \simeq 0.6l + 0.4k + 0.015t + 0.5$, i.e., actual output equals potential (as determined by the production function estimated above). This matches one of the variables that the Bank of England seems to use in judging the state of the economy. The second can be written as:

$$\begin{aligned} -\pi_t &= 0.675\left((w-p+l-g)_t + 0.005t\right) + 0.25\left(p_{£,t} - p_t\right) + 0.075\left(p_{o,t} - p_t\right) \\ &- 0.05 I_{2,t} + 1.55 \end{aligned} \tag{8.11}$$

The results confirm the discussion in the introduction to this chapter that no 'single-cause' explanation of price inflation suffices, with seven groups of forces all mattering. However, the variables for ppp deviations, excess money, national debt, and labour demand (unemployment and wages, other than through the markup) were not found to matter. Thus, fiat money has most certainly not been the only direct cause of inflation in the modern UK economy, even if commodity money (usually gold) may have played that role in the 15th–18th centuries after the imports thereof plundered from South America.

8.6 Models with expectations

The New-Keynesian Phillips curve (NKPC) includes expected future inflation as a 'feed-forward' variable to explain current inflation, often written as:

$$\Delta p_t = \underset{\geq 0}{\gamma_1} \mathsf{E}_t\left[\Delta p_{t+1} \mid \mathcal{I}_t\right] + \underset{\geq 0}{\gamma_2} \Delta p_{t-1} + \underset{\geq 0}{\gamma_3} s_t + u_t \tag{8.12}$$

where $\mathsf{E}_t\left[\Delta p_{t+1}|\mathcal{I}_t\right]$ denotes today's expected inflation one-period ahead given $\mathcal{I}_t$, the information available today, s_t denotes firms' real marginal costs and the anticipated signs of the coefficients are shown: see Galí and Gertler (1999), Galí, Gertler, and Lopez-Salido (2001) and Castle, Doornik, Hendry, and Nymoen (2014).

To make (8.12) operational, $\mathsf{E}_t[\Delta p_{t+1}|\mathcal{I}_t]$ is replaced by the actual outcome plus an error v_{t+1}:

$$\mathsf{E}_t\left[\Delta p_{t+1} \mid \mathcal{I}_t\right] = \Delta p_{t+1} + v_{t+1} \tag{8.13}$$

where by taking expectations on both sides of (8.13):

$$\mathsf{E}_t\left[\Delta p_{t+1}|\mathcal{I}_t\right] = \mathsf{E}_t\left[\Delta p_{t+1}|\mathcal{I}_t\right] + \mathsf{E}_t[v_{t+1} \mid \mathcal{I}_t] \tag{8.14}$$

so $\mathsf{E}_t[v_{t+1} \mid \mathcal{I}_t] = 0$, and hence v_{t+1} must be unpredictable from available information. Then, substituting (8.13) into (8.12):

$$\Delta p_t = \beta_1 \Delta p_{t+1} + \beta_2 \Delta p_{t-1} + \beta_3 s_t + \epsilon_t \tag{8.15}$$

where $\epsilon_t \sim \mathsf{D}[0, \sigma_\epsilon^2]$. Since v_{t+1} in (8.13) is not independent of Δp_{t+1}, neither is ϵ_t in (8.15), so estimation requires a set of 'exogenous' variables z_t as instruments (see Sections 4.8.3 and 6.5).

In the absence of good data on real marginal costs (see Section 1.8), we use real unit labour costs, or the wage share, $(w-p-g+l)_t$ for s_t, treated as endogenous. The estimates of (8.15) using as additional instruments $(w-p-g+l)_{t-1}$, $(w-p-g+l)_{t-2}$, $\Delta(g-l)_{t-1}$, $\Delta(g-l)_{t-2}$, $R_{L,t-1}$, and $R_{L,t-2}$, based on Chapter 6, were:

$$\begin{aligned}
\Delta p_t \quad = \quad & \underset{(0.09)}{0.65}\ \widehat{\Delta p}_{t+1} + \underset{(0.06)}{0.023}\ \widehat{(w-p-g+l)}_t + \underset{(0.08)}{0.50}\ \Delta p_{t-1} \\
& - \underset{(0.07)}{0.11}\ \Delta p_{t-2} - \underset{(0.11)}{0.05}
\end{aligned} \tag{8.16}$$

$$\begin{aligned}
\widehat{\sigma} &= 2.89\% \quad \mathsf{F}_{\mathsf{ar}}(2,141) = 1367.2^{**} \quad \mathsf{F}_{\mathsf{arch}}(1,146) = 1.78 \\
\chi^2_{\mathsf{nd}}(2) &= 70.54^{**} \quad \chi^2_{\mathsf{Sar}}(4) = 8.95 \quad \mathsf{F}_{\mathsf{het}}(8,139) = 5.34^{**}
\end{aligned}$$

All the anticipated signs are found, although the coefficients of the inflation variables add to more than unity with $(w-p-g+l)_t$ insignificant, and both the normality and heteroskedasticity tests strongly reject, probably from unmodelled location shifts. The one new statistic reported, $\chi^2_{\mathsf{Sar}}(k)$ is a test of the validity of the instruments (see Sargan, 1964). Overall, such estimates are similar to those reported for more recent quarterly data and other countries.

To check for location shifts, SIS was applied at 0.5%, retaining all the regressors in (8.16), and found 18 significant step indicators, denoted $\{S_{i,t}\}$, leading to:

$$\begin{aligned}
\Delta p_t \quad = \quad & -\underset{(0.06)}{0.034}\ \widehat{\Delta p}_{t+1} + \underset{(0.09)}{0.37}\ \widehat{(w-p-g+l)}_t + \underset{(0.05)}{0.28}\ \Delta p_{t-1} \\
& - \underset{(0.05)}{0.10}\ \Delta p_{t-2} - \underset{(0.17)}{0.68} + \{S_{i,t}\}
\end{aligned} \tag{8.17}$$

$$\begin{aligned}
\widehat{\sigma} &= 1.74\% \quad \mathsf{F}_{\mathsf{ar}}(2,123) = 2.75 \quad \mathsf{F}_{\mathsf{arch}}(1,146) = 2.72 \\
\chi^2_{\mathsf{nd}}(2) &= 6.81^{*} \quad \chi^2_{\mathsf{Sar}}(18) = 33.56^{*} \quad \mathsf{F}_{\mathsf{het}}(24,121) = 0.92
\end{aligned}$$

Now $\widehat{\gamma}_1$ is negative and insignificant, $(w-p-g+l)_t$ is highly significant, and inflation inertia has disappeared. Most of the mis-specification tests are also insignificant. The rejection at 5% on $\chi^2_{\mathsf{Sar}}(18)$ is due to a step indicator at 1974 being omitted from the model but included in the instruments, and adding it back to the model produces $\chi^2_{\mathsf{Sar}}(17) = 26.61$. The remaining non-normality reflects the stringent significance level of 0.5% used when selecting by SIS, and the omission from (8.17) of the important regressors in (8.10) where $\widehat{\sigma} = 1.1\%$.

As Castle, Doornik, Hendry, and Nymoen (2014) found in their analysis of NKPC models (based on IIS), the apparent significance of Δp_{t+1} in equations like (8.16) is due to the future value acting as a proxy for the unmodelled shifts. It would have taken remarkable prescience for anyone in 1914 to have anticipated the dramatically higher inflation of 1915, or even in 1916 anticipated a doubling in 1917; and they would have had to anticipate price controls restraining inflation during the Second World War.

8.7 Chapter 8 key points

(A) Viable empirical models of price inflation can also be developed.
(B) Δp_t depended on most theory-relevant domestic and foreign variables, yet the model still had many large shifts.
(C) The empirical analyses started from GUMs with all the theory-suggested variables, their dynamics, shifts and trends.
(D) Location shifts were reasonably captured by impulse and step-indicators.
(E) The price inflation model had two equilibrium corrections, one from the excess demand for goods and services, and the other from the price markup over home and imported costs.
(F) Despite its emphasis in some theories, there was little direct role for unemployment in the price inflation process.

8.8 Task 29: Forecasting with an intercept correction

We first need to create an indicator with the value unity at and after the forecast origin, but zero before, denoted $1_{\{t \geq T\}}$, to act as the intercept correction (IC). Estimating an equation with such an indicator will ensure a perfect fit at the forecast origin, so deliver an IC coefficient almost equal to the last in-sample residual.

Consider the simple equation:

$$y_t = \beta_0 + \phi 1_{\{t \geq T\}} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}[0, 1] \tag{8.18}$$

First, when no indicator is included, the model is simply:

$$y_t = \beta_0 + \nu_t \tag{8.19}$$

so when $\phi \neq 0$, estimating β_0 leads to:

$$\widetilde{\beta}_0 = T^{-1} \sum_{t=1}^{T} y_t = \beta_0 + T^{-1}\phi + T^{-1} \sum_{t=1}^{T} \epsilon_t = \beta_0 + T^{-1}\phi + \bar{\epsilon} \tag{8.20}$$

Letting $\widetilde{\epsilon}_t = y_t - \widetilde{y}_t$ where $\widetilde{y}_t = \widetilde{\beta}_0$:

$$\widetilde{\epsilon}_t = \left(\beta_0 - \widetilde{\beta}_0\right) + \phi 1_{\{t \geq T\}} + \epsilon_t = \epsilon_t - \bar{\epsilon} + \left(1_{\{t \geq T\}} - T^{-1}\right)\phi$$

and hence the last residual is:

$$\widetilde{\epsilon}_T = \epsilon_T - \bar{\epsilon} + \left(1 - T^{-1}\right)\phi \tag{8.21}$$

leading to a forecast error for y_{T+1} of $\widetilde{\epsilon}_{T+1} = y_{T+1} - \widetilde{y}_{T+1|T}$, so:

$$\widetilde{\epsilon}_{T+1} = \left(\beta_0 - \widetilde{\beta}_0\right) + \phi + \epsilon_{T+1} = \epsilon_{T+1} - \bar{\epsilon} + \left(1 - T^{-1}\right)\phi \tag{8.22}$$

When ϕ is large, substantial forecast errors can result.

However, let $\Delta\widetilde{\epsilon}_{T+1} = \widetilde{\epsilon}_{T+1} - \widetilde{\epsilon}_T$, then from (8.21) and (8.22), notice that:

$$\Delta\widetilde{\epsilon}_{T+1} = \epsilon_{T+1} - \bar{\epsilon} + \left(1 - T^{-1}\right)\phi - \left(\epsilon_T - \bar{\epsilon} + \left(1 - T^{-1}\right)\phi\right) = \Delta\epsilon_{T+1} \tag{8.23}$$

so the shift has been eliminated. Thus, adding $\widetilde{\epsilon}_T$, the last residual, to $\widetilde{y}_{T+1|T} = \widetilde{\beta}_0$ to create the corrected forecast:

$$\widehat{y}_{T+1|T} = \widetilde{\beta}_0 + \widetilde{\epsilon}_T \tag{8.24}$$

leads to the forecast error $\widehat{\epsilon}_{T+1} = y_{T+1} - \widehat{y}_{T+1|T}$ where:

$$\widehat{\epsilon}_{T+1} = \beta_0 + \phi + \epsilon_{T+1} - \widetilde{\beta}_0 - \widetilde{\epsilon}_T \tag{8.25}$$

which from (8.21) and (8.22) cancels the shift as in (8.23) to deliver:

$$\widehat{\epsilon}_{T+1} = \epsilon_{T+1} - \epsilon_T = \Delta\epsilon_{T+1} \tag{8.26}$$

which is the difference of the error that would have been made using just $\widetilde{\beta}_0$. For large magnitude ϕ, such differencing can substantively improve forecasts.

The final step is to notice that adding an impulse indicator $1_{\{t=T\}}$ for the final in-sample observation (the forecast origin) to the model in (8.19) leads to the last in-sample residual being exactly zero (within the limits of a computer's computational accuracy). The indicator $1_{\{t=T\}}$ shortens the sample by one observation, so the in-sample parameter estimates are slightly changed by adding it. However, the coefficient of the indicator is usually close to $\widetilde{\epsilon}_T$, so implements the intercept correction. That outcome is not affected by defining the indicator $1_{\{t\geq T\}}$, which carries the correction into future periods, as was done with the real-wage forecasts in Figure 6.7.

To carry out this approach to the Task, first obtain the original forecasts for the selected model in Task 23. Return to `Model`, `Formulate`, `OK`, `OK`, then set the estimation sample to end in 2011, with 7 forecasts, `OK`. Next, `Test`, tick `Forecast...`, tick h-step forecasts with h=1 to see the RMSE. Save or rename the forecast graph.

To implement the IC for forecasting over 2005–2011 from 2004, create a dummy variables that is zero till 2003, then unity after:
IC2004 = dummy(2004,1,2011,1);
Next add IC2004 to the equation (`Formulate`, `Lags`=0, double click on IC2004), `OK`, `OK`, `OK`, and repeat the forecast analysis. Finally compare the RMSEs and forecast graphs for models with and without ICs.

To compare the residual-based approach to this Task, re-estimate the model from Task 23, click on `Test`, tick `Store Residuals etc. in Database`, `OK`, tick `Residuals` in the `Store in Database` dialog, and name the residuals (e.g., *NoICResDmp*). Look for the observation for 2004 in the last column of the database that stores *NoICResDmp*, which is the final residual. This should be −.00416472 compared to the IC coefficient of −0.00428626. Thus, the two forms of correction are close in this case.

8.9 Task 30: More on forecasting

Breaks, and especially location shifts are very difficult to forecast. Indeed, even after the occurrence of outliers and shifts for UK price inflation, it was difficult to explain them using a wide range of economically-relevant variables and functions thereof, signified by the need for the many impulse, $I_{d,t}$, and step indicators, $\{S_{i,t}\}$, found. The WWI high inflation and post-war crash are primarily accounted for by indicators, and involved the major institutional change of the temporary introduction of wage indexation, so there is little hope of 'forecasting' that period even *ex post*. However, the 1970's high inflation offers a possibility despite the many 'wage and price controls' tried by successive governments.

First, the combined indicator, $I_{d,t}$, has to be revised to eliminate the impulses for 1971 onwards. Then, for the model in (8.10), select the sample 1875–1982 with 12 'forecasts', over 1971–1982. The RMSE= 0.044, so is more than three times as large as $\widehat{\sigma} = 0.013$, with a significant non-constancy test $\mathsf{F}_{\mathsf{Chow}}(12, 86) = 7.59^{**}$, with many large forecast errors, most of which are positive so the outcomes are under-predicted. Figure 8.7 panels a and b show the forecasts, $\Delta\widehat{p}_{T+h+1|T+h}$, with 95% error bars and forecast errors respectively.

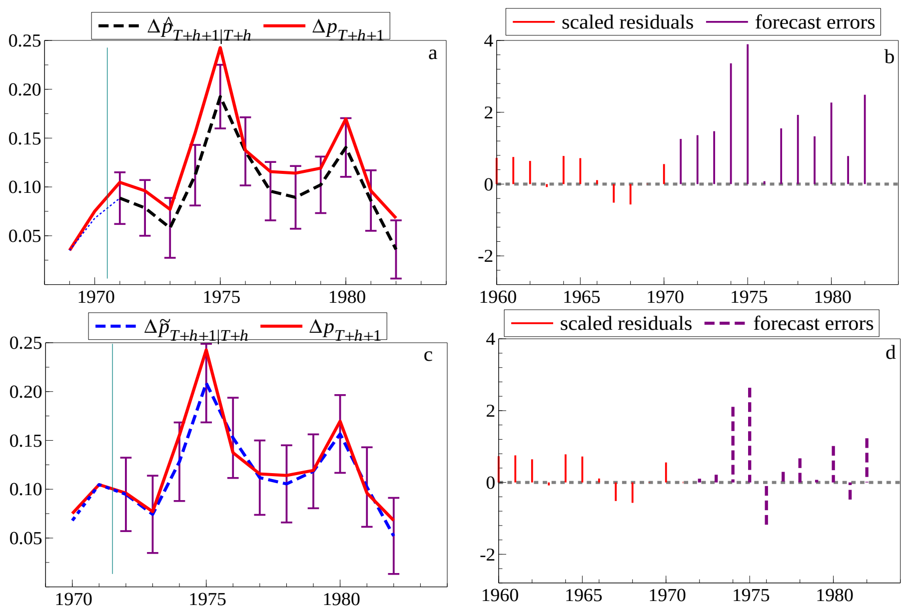


Figure 8.7
1970's price inflation 'forecasts' with and without intercept correction

Next, create an intercept correction indicator with zero till 1970 and unity from 1971–1982, as explained in Task 29 Section 8.8. Add it to the equation, moving one period forward to estimate up to and including 1971, and forecast the remaining 11 years (1-step ahead each time, but without updating). The improvement is marked: RMSE= 0.016 as against $\widehat{\sigma} = 0.013$, and $\mathsf{F}_{\mathsf{Chow}}(11, 86) = 0.83$, so now the model forecasts better than it fits over that turbulent period. Figure 8.7 illustrates the results graphically in panels c and d. The forecast errors from $\Delta\widetilde{p}_{T+h+1|T+h}$ are much smaller, and not all are positive, with no outcomes outside the 95% error bars.

8.10 Task 31: Cointegration simulations

We begin with a short analysis of cointegration in a simple setting. Consider the following model over $t = 1, 2, \ldots, T$:

$$\begin{aligned} y_t &= \mu_0 + \lambda x_t + e_t \ \text{ with } \ e_t \sim \mathsf{IN}[0, \sigma_e^2] \\ z_t &= \mu_1 + x_t + \nu_t \ \text{ and } \ \nu_t \sim \mathsf{IN}[0, \sigma_\nu^2] \end{aligned} \tag{8.27}$$

when:

$$x_t = x_{t-1} + \beta + \epsilon_t \ \text{ where } \ \epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2] \tag{8.28}$$

and x_0 is a fixed number. Then x_t can be expressed in terms of current and past values of ϵs, x_0, and β:

$$x_t = x_{t-1} + \beta + \epsilon_t = x_{t-2} + 2\beta + \epsilon_t + \epsilon_{t-1} = \cdots = x_0 + \beta t + \sum_{s=1}^{t} \epsilon_s$$

so $\{x_t\}$ is a random walk with drift β, which is also a stochastic trend at rate β. Since y_t and z_t depend on x_t, they are also both I(1).

Cointegration requires us to obtain a relationship between y_t and z_t that does not depend on x_t, which can be done by subtracting λ times z_t from y_t:

$$\begin{aligned} y_t - \lambda z_t &= \mu_0 + \lambda x_t + e_t - \lambda\mu_1 - \lambda x_t - \lambda\nu_t \\ &= (\mu_0 - \lambda\mu_1) + e_t - \lambda\nu_t \end{aligned}$$

which only depends on I(0) errors and not on x_t, so y_t and z_t are cointegrated.

It is possible to prove that such a relationship between y_t and z_t is unique by considering any other value δ than λ so that:

$$\begin{aligned} y_t - \delta z_t &= \mu_0 \lambda x_t + e_t - \delta\mu_1 - \delta x_t - \delta\nu_t \\ &= (\mu_0 - \delta\mu_1) + (\lambda - \delta) x_t + e_t - \delta\nu_t \end{aligned}$$

which depends on $\{x_t\}$ when $\delta \neq \lambda$.

Task 15 Section 4.15 explained the basics of Monte Carlo simulations. The experiment here involves creating a DGP with two I(1) cointegrated series, y_t and z_t, where z_t has a large location shift, so requires the `Advanced Experiment` option. Then we will simulate estimation and inference in an AD(1,1) model.

Select `Model`, and change the `Category` to `Monte Carlo`, and then the `Model class` to the option `Advanced Experiment using PcNaive & Ox Professional`. The experimental design is stored in PcNaive_CI1.ox, and you will need *Ox Professional* to run that file (this is essentially the same file for the next experiment in Task 31: only the model differs). The DGP is:

$$\begin{aligned} y_t &= \beta_1 y_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \epsilon_t \ \text{ where } \ \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \\ z_t &= z_{t-1} + \lambda 1_{\{T_1 \le t \le T_2\}} + e_t \ \text{ where } \ e_t \sim \mathsf{IN}\left[0, \sigma_e^2\right] \end{aligned} \tag{8.29}$$

The selected DGP parameter values are $\beta_1 = 0.8$, $\beta_2 = 0.5$, $\beta_3 = -0.3$, with $\sigma_\epsilon = \sigma_e = 1$, $T = 100$, $\lambda = 0.5$, $T_1 = 50$, $T_2 = 70$, $M = 10000$ replications. The model is the first equation in (8.29), which can be expressed as the EqCM:

$$\Delta y_t = \beta_2 \Delta z_t - (1 - \beta_1)(y_{t-1} - \kappa z_{t-1}) + \epsilon_t$$

where $\kappa = (\beta_2 + \beta_3)/(1 - \beta_1)$ (equal to unity for the values here), but that information is not used in the simulations.

There is considerable graphical output during the simulations (shown in Figure 8.8 below), and at their termination, output will be stored in PcNaive_CI1.out. The key aspects here are the results for rejection frequencies (shown in Table 8.1) and moments of estimates (shown in Table 8.2). In the former, for example, t-Ya_1 in the output denotes the t-test for the null hypothesis that the coefficient β_1 of y_{t-1} is zero, and so on: thus, that is rejected 100% of the time at both 1% and (hence also) 5%. Similarly, $\mathsf{H}_0 : \beta_2 = 0$ and $\mathsf{H}_0 : \beta_3 = 0$ are rejected 99% and 41% at $\alpha = 0.01$, with the last rising to 65% at $\alpha = 0.05$.

	1%	5%
t-Ya_1	1.0000	1.0000
t-Yb	0.98710	0.99730
t-Yb_1	0.40560	0.65340
AR1	0.0095000	0.050800
DW	0.0021000	0.021300
Hetero	0.015700	0.053100
H_z(p<=0|n+1)	0.99990	1.0000
[ASE]	0.00099499	0.0021794

Table 8.1
Monte Carlo rejection frequencies

Next, AR1, DW, and Hetero show the rejections on those mis-specification tests (for first-order residual autocorrelation, the Durbin–Watson test, and for residual heteroskedasticity–the first and third being the tests used above), which are roughly equal to the nominal significance, except that DW is not appropriate in AD(1,1) models so under-rejects (see Nerlove and Wallis, 1966). The second bottom row shows the rejection of the null of no cointegration, which correctly does so 100% of the time. The last row, [ASE], shows the value of $\sqrt{\alpha \times (1-\alpha)/M}$ which is the standard error under the null of estimating a probability of α from M trials.

	mean	MCSD
Ya_1	0.77262	0.060455
Yb	0.49824	0.10024
Yb_1	-0.27265	0.11450
ESE[Ya_1]	0.057000	0.0079860
ESE[Yb]	0.099197	0.010176
ESE[Yb_1]	0.11503	0.010649
sigma^ 2	0.99785	0.14384
R^ 2	0.95951	0.045363

Table 8.2
Monte Carlo moments of estimates

The means and Monte Carlo Standard Deviations (MCSDs) of the estimates are shown in Table 8.2, as well as their Estimated Standard Errors (ESEs). The MCSD are calculated as shown in (4.52), so represent the actual variability in the coefficient estimates generated by the simulation. In contrast, the mean ESEs, which are the 'standard errors' reported below coefficient estimates in empirical regressions, are based on the square root of the right-hand side formula in (4.16), denoted SE[$\widehat{\beta}_i$] above. The derivation of (4.16) made many assumptions (including IN errors, constant parameters, accurate data, etc.), and only when all of those assumptions are satisfied will the ESE equal the corresponding MCSD, which happens here (despite the data being I(1)). For example, the MCSD for Yb_1 is 0.1145 and the corresponding ESE is 0.11503.

Here, the MCSDs are about 10% of the mean ESEs, so the finding would be reported as (e.g.) ESE[Yb_1]=0.1150±0.0002 for a 95% interval, so that is very accurately determined.

Finally, sigmaˆ 2 and Rˆ 2 are the respective outcomes for $\widehat{\sigma}^2_\epsilon$ (where the DGP value is unity) and R^2, where the very high value reflects the data being I(1). Returning to the mean estimated parameter values these are close to their respective DGP values of $\beta_1 = 0.8$, $\beta_2 = 0.5$, $\beta_3 = -0.3$, so least-squares estimation of the AD(1,1) worked well even for this cointegrated setting.

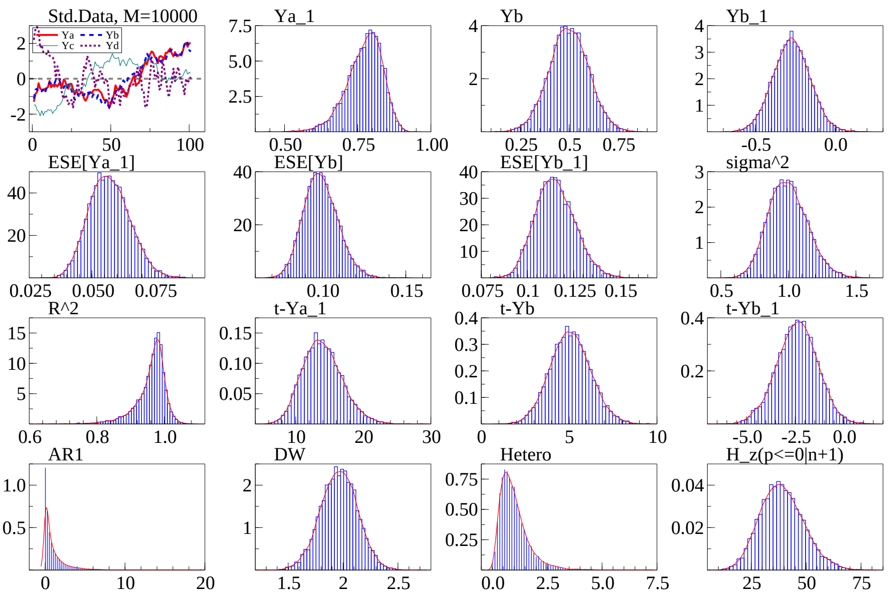


Figure 8.8
Cointegration simulations graphical output

The simulation graphs in Figure 8.8 report the output at the last replication, so record the last draw of all the data series in the top-left plot, the histograms and empirical density estimates in the remainder of the top row, then the densities of the three ESEs (from which the MCSDs just discussed are calculated), the densities of sigmaˆ 2 and Rˆ 2, then the densities of the t-tests of the null that each β_i is zero (from which the rejection frequencies in Table 8.1 are calculated), and the last row shows the densities of the test statistics, the first three under the null and the last for cointegration under the alternative.

8.11 Task 32: Understanding integration and cointegration

The final Task for this chapter returns to Section 2.6 to investigate why so many economic time series are integrated. Imagine an economy with just consumption, c_t, and income, y_t, where these two variables are connected by:

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 c_{t-1} + \epsilon_{1,t} \text{ where } \epsilon_{1,t} \sim \mathsf{IN}\left[0, \sigma_1^2\right] \tag{8.30}$$

with $|\beta_3| < 1$ and:

$$y_t = \lambda_0 + \lambda_1 c_{t-1} + \lambda_2 y_{t-1} + \epsilon_{2,t} \text{ with } \epsilon_{2,t} \sim \mathsf{IN}\left[0, \sigma_2^2\right] \tag{8.31}$$

where $|\lambda_2| < 1$. These equations can be re-arranged into a more interpretable form by subtracting their lagged values from the left-hand side variables to express them as changes, Δc_t and Δy_t, then combining the remaining levels terms as 'feedbacks'. For example:

$$\begin{aligned}
\Delta c_t &= \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \left(\beta_3 - 1\right) c_{t-1} + \epsilon_{1,t} \\
&= \beta_0 + \beta_1 \Delta y_t + \left(\beta_1 + \beta_2\right) y_{t-1} + \left(\beta_3 - 1\right) c_{t-1} + \epsilon_{1,t} \\
&= \beta_0 + \beta_1 \Delta y_t + \left(1 - \beta_3\right)\left(\kappa y_{t-1} - c_{t-1}\right) + \epsilon_{1,t}
\end{aligned}$$

where $\kappa = \left(\beta_1 + \beta_2\right) / \left(1 - \beta_3\right)$. When $\kappa = 1$:

$$\Delta c_t = \beta_1 \Delta y_t + \gamma(y_{t-1} - c_{t-1} - \mu_0) + \epsilon_{1,t} \tag{8.32}$$

where $y_t - c_t$ is saving, $1 > \left(1 - \beta_3\right) = \gamma > 0$ and $\mu_0 = \beta_0 / \left(\beta_3 - 1\right)$. Thus, the change in consumption responds by β_1 to a change in income, and by γ to the previous deviation of savings from μ_0. Similarly:

$$\Delta y_t = \delta(y_{t-1} - \theta c_{t-1} - \mu_1) + \epsilon_{2,t} \tag{8.33}$$

where $\delta = \lambda_2 - 1 < 0$, $\theta = \lambda_1/(1 - \lambda_2)$ and $\mu_1 = \lambda_0/(1 - \lambda_2)$. This derivation also establishes that equilibrium-correction formulations are equally applicable to $\mathsf{I}(0)$ processes.

In a non-stochastic equilibrium where all changed has ceased, denoting the resulting outcomes by a superscript *, so $\mathsf{E}[\Delta c_t^*] = \mathsf{E}[\Delta y_t^*] = 0$, then (8.32) and (8.33) respectively imply that $y^* = c^* + \mu_0$ and $y^* = \theta c^* + \mu_1$, which are two equations in the two unknowns c^* and y^*, with the unique solution that $c^* = (\mu_1 - \mu_0)/(1 - \theta)$ and $y^* = \mu_0 + (\mu_1 - \mu_0)/(1 - \theta)$ when $1 > \theta > 0$. This economy converges on that point, and so c_t and y_t are stationary.

Such an outcome no longer holds if $\theta = 1$, imposing $\mu_1 = \mu_0$ for consistency. When the two equations respond to just one equilibrium correction, $(y_{t-1} - c_{t-1} - \mu_0)$, the variables c_t and y_t are integrated of order 1, $\mathsf{I}(1)$, and cointegrated. Thus, the economy now moves along an equilibrium trajectory with a constant equilibrium savings rate, but changing levels of c_t and y_t. Thinking of $(y_{t-1} - c_{t-1} - \mu_0)$ as the decision-making variable, and c_t and y_t as the decisions, then the general result is that data are integrated when there are fewer decision-making variables than decisions. If all economic variables are eventually connected (like a water bed, where a push anywhere disturbs the bed everywhere), then all economic variables are integrated, and the number of decision-making variables like $(y_{t-1} - c_{t-1} - \mu_0)$ in (8.32) determine the number of separate cointegrated relationships. This fundamental result is due to Clive Granger—see Robert Engle and Granger (1987): Hendry (2004) reviews the history and offers further explanation.

8.12 Chapter 8 exercises

1. Create and execute the following Monte Carlo study using the setup instructions in Task 15 Section 4.15, but selecting the `Model class` as `Static Experiment using PcNaive`. The DGP is:

$$y_t = \beta_0 + \beta_1 z_{1,t} + \beta_2 z_{2,t} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2\right] \tag{8.34}$$

   where $z_{i,t} \sim \mathsf{IN}[0, 1]$ and $\sigma^2 = 1$. The aim is to compare the impact on the distribution of $\widehat{\beta}_1$ when omitting the regressor $z_{2,t}$ as against including it, once when it is orthogonal, and once when it is highly correlated with $z_{1,t}$. Set $\beta_1 = 1$ (the default) and $\beta_2 = 2$, noting $\beta_0 = 0$, using 1000 replications. Accept all the other default options, but tick all the `Monte Carlo Output` boxes except `AR(1) test, DW`, and all the `Live Graphics` options, choosing a plot frequency of 100. Two models need to be formulated, one including and the other excluding $z_{2,t}$, denoted 'Zb' in the *PcNaive* program. Excluding 'Zb' is the default so run that experiment first, rename the output graph, then rerun with the `Zb` box ticked.
2. Check that the data generated in both experiments is identical, then change the second set of estimated densities for $\widehat{\beta}_1$, $\mathsf{SE}[\widehat{\beta}_1]$, $\widehat{\sigma}^2$ and $\mathsf{t}_{\widehat{\beta}_1}$ by light grey shading inside the histograms, where the first (mis-specified) set are denoted $\widetilde{\beta}_1$, $\mathsf{SE}[\widetilde{\beta}_1]$, $\widetilde{\sigma}^2$ and $\mathsf{t}_{\widetilde{\beta}_1}$.
3. Paste these second set of density plots on the corresponding first set, only retaining those four plots.
4. Return to `Formulate`, change the `Correlation between Za and Zb` to 0.95 (denoted ρ below), exclude 'Zb' from the model again and rerun, renaming the graph. Repeat with 'Zb' included, then follow the same procedures as *iii*). The outcome, combined in one graph, should be like Figure 8.9.
5. Derive the expected value $\mathsf{E}[\widetilde{\beta}_1]$ in both cases. Can you derive the variance $\mathsf{Var}[\widetilde{\beta}_1]$ in both cases?
6. If $\rho = \mathsf{corr}[z_{1,t}, z_{2,t}]$ remained constant over a forecast period, would the omission of $z_{2,t}$ matter in either case? Would your answer change if ρ shifted over the forecast horizon?

Panels *a–d* in Figure 8.9 report outcomes for the orthogonal $z_{2,t}$, and *e–h* are for when $\mathsf{corr}[z_{1,t}, z_{2,t}] = 0.95$. You should be able to discern advantages in both settings from avoiding a serious mis-specification. When $\mathsf{corr}[z_{1,t}, z_{2,t}] = 0$, although $\widetilde{\beta}_1$ is not biased for β_1, its density is more spread out, the ESE and $\widetilde{\sigma}^2$ are much larger, and the density of the t-test is shifted towards the origin.

When the omitted variable $z_{2,t}$ is highly correlated with the included, a very different picture emerges. Now $\widetilde{\beta}_1$ is badly biased for β_1, but has a much smaller ESE, as the high correlation between the regressors greatly increases the uncertainty as to the true value of β_1 when both regressors are correctly included. This increased variance carries over to a smaller t-test distribution for $\widehat{\beta}_1$ even though $\widehat{\sigma}^2$ is smaller than $\widetilde{\sigma}^2$.

To summarize, $\widetilde{\beta}_1$ is precise, but is actually estimating the wrong parameter, namely $\beta_1 + \rho\beta_2$: that could matter greatly for economic policy, but for a constant ρ, may have only a small impact on forecasts.

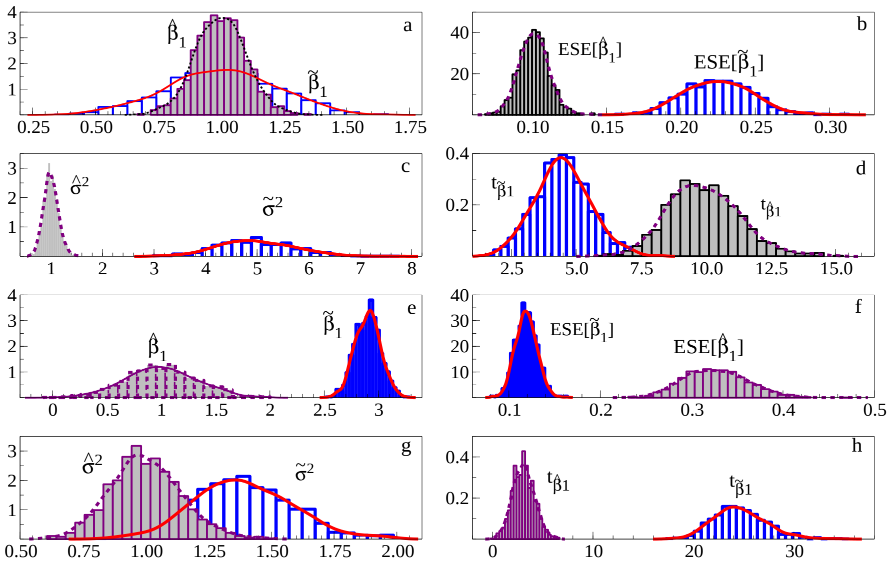


Figure 8.9
Simulation graphical output for correctly and incorrectly specified models

Chapter 9
Conclusions

Huge changes have occurred over the past 150 years in technological and financial innovations, new laws, health, demography, social reforms, wars, educational availability, policy-regime shifts, and so on. Overall, these combined have resulted in great improvements in real wages, but also in large rises in the price level, and varying levels of unemployment. Breaks and shifts were pervasive, and needed to be modelled to understand how economies work. To do so, we used differences, cointegration, impulse and step indicator saturation, co-breaking, dynamics, non-linear functions and related variables to tackle inter-dependence, evolution, abrupt shifts, and changing relationships.

Modelling methodology yields insights into how to undertake and evaluate empirical analyses, and emphasizes that all forms of non-stationarity must be modelled, including stochastic trends, measurement changes, non-linearities and location shifts. Moreover, all substantively relevant variables and their dynamics must be included, so initial general unrestricted models will be large even before tackling non-linearities and indicator saturation. The large numbers of potentially relevant variables requires computer software to choose the empirical models, a move from humans doing so that is analogous to the earlier application of computers to regression analysis replacing calculations previously done by hand. Indicator saturation can only be undertaken by a computer program–but *Autometrics* does so successfully as we have seen.

The selected models of real-wage growth and price inflation explain much of the observed movements over almost 150 years, but despite allowing for many potential influences, still needed indicator variables for the largest shifts. The unemployment rate was also subject to major location shifts, which help explain apparent non-constancies in the 'Phillips curve'. However, neither unemployment nor expectations of future inflation played direct roles in either real-wage or price-inflation equations. Money demand was difficult to model as financial innovations changed its measurements intermittently. No role was found for 'excess money' in inflation, although money growth was significant.

The teaching of empirical macro-econometrics often eschews the complications that are ever present in macroeconomic data because they are 'too difficult'. Having seen the prevalence of strong yet changing trends, abrupt shifts, high correlations between variables and with their own lags, this short introduction could not sensibly avoid tackling their joint interactions, so addressed all the main concepts and some of the basic tools. The power, flexibility and ease of use of computer software that can reduce a high-dimensional initial formulation to an interpretable, parsimonious and encompassing final selection makes it feasible to adopt such an approach to teaching, and many of the models presented above can be illustrated live in the classroom, hopefully empowering future generations of empirical modellers.

Notwithstanding the difficulties many experienced after the 2008 Financial Crisis and ensuing Great Recession, and valid worries about inequality, climate change and new pandemics, the present is one of the best ever times in history to be alive: I cannot imagine many readers wanting to go back and live a median life in 1860. Although living 150 years ago was not quite as bad as "the life of man, solitary, poor, nasty, brutish, and short" (to quote from Thomas Hobbes's *Leviathan*, 1651), about 15% of infants died in their first year of life. The vast improvements in almost every aspect of life are beautifully captured visually in http://www.ourworldindata.org/: change is the norm in them all, just as we have seen in macroeconomics.

Appendix: Data measurements

The variables used in the above analyses were defined as follows:

P_t	=	implicit deflator of GDP, (1860=1)	[6], p.836, [9]a (1993), [20] code:ABML.
G_t	=	real GDP, £million, 1985 prices	[6], p.836, [9]a (1993), [20] code:YBHH.
U_t	=	unemployment	[7], [9]c (1993), [19] code: MGSC.
$Wpop_t$	=	working population	[7], [9]c (1993), [19] code: MGSF.
$U_{r,t}$	=	$U_t/Wpop_t$ (unemployment rate)	
L_t	=	employment (= $Wpop_t - U_t$)	[4], [5]
W_t	=	average weekly earnings	[17], [18], [19] code: LNMM
K_t	=	total capital stock	[6], p.864, [9]c (1972,1979,1988,1992)
$R_{L,t}$	=	bond rate	[1], [2]
$R_{S,t}$	=	short-term interest rate	[1], [2]
M_t	=	nominal broad money, £million	[1], [2]
$R_{n,t}$	=	opportunity-cost of money measure	[3]
N_t	=	nominal National Debt, £million	[8]
E_t	=	£ effective exchange rate index	[1], [2], [10]
$P_{w,t}$	=	trade-weighted world price index	[1], [2], [10]
$P_{o,t}$	=	price index, raw materials & fuels	[11]
TU_t	=	Trades Unions membership	[6], p.137, [15], Table 6.20, [16]
B_t	=	replacement ratio	[12] code: 1.6.4/NQDNAU, [13], [14] 1989 Table 34.01, 1992 Table A2.36;
S_t	=	days lost through strikes, millions	[6], p.144 and AA,Table 6.1; [21]
NIC_t	=	National Insurance Contributions	[12] code: 11.1 / CEANAU
$W_{r,t}$	=	nominal wage rates	[5], [12], [18]
H_t	=	normal hours (from 1920)	[6], p.148, [9]
ULC_t	=	unit labour costs $(= L_t W_{r,t}/G_t)$	
I_d	=	combined indicator: see below	[22]
1_{abcd}	=	indicator equal to unity in year $abcd$	
Δx_t	=	$(x_t - x_{t-1})$ for any variable x_t	
$\Delta^2 x_t$	=	$\Delta x_t - \Delta x_{t-1}$	

Sources:

[1] Friedman and Schwartz (1982);
[2] Attfield, Demery, and Duck (1995);
[3] Ericsson, Hendry, and Prestwich (1998);
[4] Shadman-Mehta (1995) (citing Sleeman, 1981. and Thomas, 1984, as sources);
[5] Phillips (1958);
[6] Mitchell (1988);
[7] Feinstein (1972);
[8] Bank of England;
[9] Bean (from (a) *Economic Trends Annual Supplements*, (b) *Annual Abstract of Statistics*, (c) *Department of Employment Gazette*, and (d) *National Income and Expenditure*,

as well as other sources cited here);
[10] from Gavin Cameron and John Muellbauer;
[11] UN Statistical Yearbook and Christopher Gilbert;
[12] Office for National Statistics (ONS), *Blue Book*;
[13] Board of Trade (1860–1908);
[14] *Social Security Statistics* (now *Social Security and Child Support Statistics*);
[15] ONS, *Annual Abstract of Statistics*;
[16] ONS, *Labour Market Trends*;
[17] Crafts and Mills (1994);
[18] Feinstein (1990);
[19] ONS, *Labour Force Survey*;
[20] ONS, *Economic Trends Annual Supplement*;
[21] ONS, *Social Trends*;
[22] Hendry (2001).
Notes:
Average weekly wages: a measure of full-time weekly earnings for all blue collar workers, where the coverage has been extended to include more occupations, and allows for factors such as changes in the composition of the manual labour force by age, sex, and skill, and the effect of variations in remuneration under piece rates and other systems of payments, but not adjusted for time lost through part-time work, short-time, unemployment etc. A reduction in standard hours worked that was offset by a rise in hourly wage rates would not be reflected in the index. From 1855–1880, the data are from Feinstein (1990), but not revised to increase coverage. Prior to that, the data come from a number of sources on average wage rates for blue collar workers.
Nominal wage rates: hourly wage rates prior to 1946, then weekly wage rates afterwards, so the latter were standardized by dividing by normal hours. The trend rate of decline of hours is about 0.5% p.a. (based on a drop from 56 to 40 hours per week between 1913 and 1990, with an additional increase in paid holidays), and spliced to an average earnings index for the whole economy including bonuses [ONS: LNMM] from 1991 and rebased to 2000=1. The average earnings index was discontinued in 2010, and replaced with average weekly earnings. The unit labour cost measure ulc^* was adjusted accordingly.
The 'replacement ratio' is the ratio of unemployment benefits to average wages. Benefits are measured by the amount expended for poor relief in unionized industries/number of paupers (from [13]) to 1908 (data for England and Scotland). Interwar and post-WWII data on value of benefits from [14], Table 34.01 (1989), Table A2.36 (1992) and [12] total government benefit expenditure/population. Data are spliced over 1909-1919 and 1939-1945. The unemployment data were based on Trades Unions members till 1945. The unemployment rate is measured as a fraction. Hours lost through strikes, p.a., are only available from 1890.
For GDP, code:YBHH was at 2005 prices, so scaled to 1985.
Exchange rate is the annual £/$ rate till 1954, then an annual aggregate of quarterly data on the trade-weighted effective exchange rate, spliced to the £/$ rate in 1955. World prices were US prices till 1954, then a trade-weighted annual aggregate of quarterly data on the corresponding *PPP* values, from which the price data were derived and spliced to US prices in 1955. Commodity prices are a composite price index for raw materials and fuels. Revised data from 1868–1933 were kindly provided by Christopher Gilbert, and spliced to the index in Hendry (2001).

The short-term interest rate $R_{S,t}$ is the three-month treasury bill rate, fraction p.a., and the long-term interest rate $R_{L,t}$ is the long-term bond rate, fraction p.a.
Nominal money is a spliced series of broad money measures, using M2 ('old definition') till 1968; M3 from 1963 though 1987; and M4 (adjusted for definitional breaks) from 1982 onwards. The corresponding measure of the opportunity cost of holding money, $R_n = R_S \times (H^a/M^a)$ where H is 'high-powered money', which is non-interest bearing, and the superscript ^{a} denotes that the actual (rather than rescaled after splicing) values were used: see Ericsson, Hendry, and Prestwich (1998).
Hendry (2001) found 22 indicators were needed to remove large outliers given by:

$$I_b = \begin{cases} 1 & \text{for } 1915,\ 1917,\ 1919 \\ -1 & \text{for } 1921,\ 1922 \end{cases}$$

$$I_l = \begin{cases} 1 & \text{for } 1916,\ 1918,\ 1920,\ 1975 \\ -1 & \text{for } 1943,\ 1945,\ 1973 \end{cases}$$

$$I_m = \begin{cases} 1 & \text{for } 1880,\ 1900,\ 1939,\ 1940,\ 1970,\ 1971,\ 1980 \\ -1 & \text{for } 1881,\ 1942,\ 1944. \end{cases}$$

where I_d combined these with weights of 3, 1.5, and 1.
Hendry and Ericsson (1991) and Hendry (2001) provide detailed discussions about many of these series.

References

Allen, R. C. (2009). *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press. [9, 144]

Allen, R. C. (2011). *Global Economic History: A Very Short Introduction*. Oxford: Oxford University Press. [2, 9]

Attfield, C. L. F., D. Demery, and N. W. Duck (1995). Estimating the UK demand for money function: A test of two approaches. Mimeo, Economics Department, University of Bristol. [163]

Baba, Y., D. F. Hendry, and R. M. Starr (1992). The demand for M1 in the U.S.A., 1960–1988. *Review of Economic Studies 59*, 25–61. [124]

Banerjee, A., J. J. Dolado, J. W. Galbraith, and D. F. Hendry (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press. [91]

Barro, R. J. (2009). Rare disasters, asset prices and welfare costs. *American Economic Review 99*, 243–264. [7]

Beveridge, W. (1942). *Social Insurance and Allied Services*. London: HMSO. [34]

Beveridge, W. (1945). *Full Employment in a Free Society*. New York: WW Norton. [34]

Bowley, A. L. (1937). *Wages and Income in the United Kingdom since 1860*. Cambridge: Cambridge University Press. [2]

Boyer, G. R. and T. J. Hatton (2002). New estimates of British unemployment, 1870–1913. *Journal of Economic History 62*, 643–675. [52]

Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*, 239–246. [95]

Castle, J. L., J. A. Doornik, D. F. Hendry, and R. Nymoen (2014). Mis-specification testing: Non-invariance of expectations models of inflation. *Econometric Reviews 33*, 553–574. [111, 150, 151]

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3(2)*, 240–264. [96, 105]

Castle, J. L. and D. F. Hendry (2009). The long-run determinants of UK wages, 1860–2004. *Journal of Macroeconomics 31*, 5–28. [108, 109]

Castle, J. L. and D. F. Hendry (2010). A low-dimension portmanteau test for non-linearity. *Journal of Econometrics 158*, 231–245. [39, 100, 112, 120, 147]

Castle, J. L. and D. F. Hendry (2014). Semi-automatic non-linear model selection. In N. Haldrup, M. Meitz, and P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 163–197. Oxford: Oxford University Press. [111, 112, 113, 114]

Castle, J. L. and N. Shephard (Eds.) (2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press. [168, 170]

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591–605. [101]

Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. [115]

Cobb, C. W. and P. H. Douglas (1928). A theory of production. *American Economic Review 18 (Supplement)*, 139–165. [10]

Coghlan, R. T. (1978). A transactions demand for money. *Bank of England Quarterly Bulletin 18*, 48–60. [124]

Crafts, N. F. R. (2002). Is economic growth good for us? Public lecture, Royal Economic Society, Glasgow, Manchester and London. [7]

Crafts, N. F. R. and T. C. Mills (1994). Trends in real wages in Britain, 1750-1913. *Explorations in Economic History 31*, 176–194. [164]

de Brouwer, G. and N. R. Ericsson (1998). Modelling inflation in Australia. *Journal of Business and Economic Statistics 16*, 433–449. [104]

Dicks-Mireaux, L. A. and J. C. R. Dow (1959). The determinants of wage inflation: United Kingdom, 1946–1956. *Journal of the Royal Statistical Society A, 122*, 145–84. [104]

Doornik, J. A. (2007). *Object-Oriented Matrix Programming using Ox* (6th ed.). London: Timberlake Consultants Press. [83]

Doornik, J. A. (2009). Autometrics. See Castle and Shephard (2009), pp. 88–121. [94]

Doornik, J. A. (2013a). *Object-Oriented Matrix Programming using Ox* (7th ed.). London: Timberlake Consultants Press. [xi]

Doornik, J. A. (2013b). *OxMetrics: An Interface to Empirical Modelling* (7th ed.). London: Timberlake Consultants Press. [xi, 20]

Doornik, J. A. and H. Hansen (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics 70*, 927–939. [79]

Doornik, J. A. and D. F. Hendry (2013). *Empirical Econometric Modelling using PcGive: Volume I.* (7th ed.). London: Timberlake Consultants Press. [xi, 20]

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007. [79]

Engle, R. F. and C. W. J. Granger (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–276. [157]

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304. [68, 71]

Ericsson, N. R., D. F. Hendry, and K. M. Prestwich (1998). The demand for broad money in the United Kingdom, 1878–1993. *Scandinavian Journal of Economics 100*, 289–324. [18, 125, 126, 143, 163, 165]

Ericsson, N. R. and J. G. MacKinnon (2002). Distributions of error correction tests for cointegration. *Econometrics Journal 5*, 285–318. [99]

Escribano, A. (2004). Nonlinear error correction: The case of money demand in the UK (1878–2000). *Macroeconomic Dynamics 8*, 76–116. [128, 143]

Feinstein, C. H. (1972). *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge: Cambridge University Press. [163]

Feinstein, C. H. (1990). New estimates of average earnings in the UK, 1880-1913. *Economic History Review 43*, 595–632. [164]

Friedman, M. (1977). Nobel lecture: Inflation and unemployment. *Journal of Political Economy 85*, 451–472. [104]

Friedman, M. and A. J. Schwartz (1963). *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press. [124]

Friedman, M. and A. J. Schwartz (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. Chicago: University of Chicago Press. [124, 131, 163]

Galí, J. and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics 44*, 195–222. [150]

Galí, J., M. Gertler, and J. D. Lopez-Salido (2001). European inflation dynamics. *European Economic Review 45*, 1237–1270. [150]

Garraty, J. A. (1979). *Unemployment in History*. London: Harper. [87]

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica 46*, 1303–1313. [79]

Godley, W. A. H. and W. D. Nordhaus (1972). Pricing in the trade cycle. *Economic Journal 82*, 853–882. [104]

Goldfeld, S. M. (1976). The case of the missing money. *Brookings Papers on Economic Activity 6*, 683–730. [124]

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press. [39]

Hacche, G. (1974). The demand for money in the United Kingdom: Experience since 1971. *Bank of England Quarterly Bulletin 14*, 284–305. [124]

Hendry, D. F. (1979). Predictive failure and econometric modelling in macroeconomics: The transactions demand for money. In P. Ormerod (Ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. [124]

Hendry, D. F. (1980). Econometrics: Alchemy or science? *Economica 47*, 387–406. [144]

Hendry, D. F. (1984). Monte Carlo experimentation in econometrics. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 16, pp. 937–976. Amsterdam: North-Holland. [82]

Hendry, D. F. (1995a). *Dynamic Econometrics*. Oxford: Oxford University Press. [147]

Hendry, D. F. (1995b). On the interactions of unit roots and exogeneity. *Econometric Reviews 14*, 383–419. [71]

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review 27*, 401–422. [146]

Hendry, D. F. (2001). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics 16*, 255–275. [2, 89, 128, 140, 142, 164, 165]

Hendry, D. F. (2004). The Nobel Memorial Prize for Clive W.J. Granger. *Scandinavian Journal of Economics 106*, 187–213. [157]

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills and K. D. Patterson (Eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan. [70]

Hendry, D. F. and N. R. Ericsson (1991). An econometric analysis of UK money demand in 'Monetary Trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz. *American Economic Review 81*, 8–38. [124, 127, 130, 143, 165]

Hendry, D. F., S. Johansen, and C. Santos (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 33*, 317–335. Erratum, 337–339. [94]

Hendry, D. F. and G. E. Mizon (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal 88*, 549–563. [124]

Hendry, D. F. and B. Nielsen (2007). *Econometric Modeling: A Likelihood Approach*. Princeton: Princeton University Press. [xi, 48, 62, 78]

Hendry, D. F. and B. Nielsen (2010). A modern approach to teaching econometrics. *European Journal of Pure and Applied Mathematics 3*, 347–369. [xi, 94]

Hendry, D. F. and C. Santos (2005). Regression models with data-based indicator variables. *Oxford Bulletin of Economics and Statistics 67*, 571–595. [95]

Houthakker, H. S. (1956). The Pareto distribution and the Cobb-Douglas production function in activity analysis. *Review of Economic Studies 23*, 27–31. [10]

Hume, D. (1752). *Political Discourses*. Edinburgh. [124]

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. See Castle and Shephard (2009), pp. 1–36. [94]

Johnson, P. (1988). Conspicuous consumption and working-class culture in Late-Victorian and Edwardian Britain. *Transactions of the Royal Historical Society 38*, 27–42. [104]

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan. [10, 104]

Layard, R., S. J. Nickell, and R. Jackman (1991). *Unemployment, Macroeconomic Performance and the Labour Market*. Oxford University Press. [104]

Mackenzie, W. A. (1921). Changes in the standard of living in the United Kingdom, 1860–1914. *Economica 3*, 211–230. [103, 104]

Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill. [39]

Marshall, A. (1926). *Official Papers*. London: Macmillan and Company. [6, 26, 124]

Milbourne, R. (1983). Optimal money holding under uncertainty. *International Economic Review 24*, 685–698. [128]

Miller, M. H. and D. Orr (1966). A model of the demand for money by firms. *Quarterly Journal of Economics 80*, 735–759. [128]

Mitchell, B. R. (1988). *British Historical Statistics*. Cambridge: Cambridge University Press. [163]

Mizon, G. E. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics 69*, 267–288. [80]

Nerlove, M. and K. F. Wallis (1966). Use of the Durbin-Watson statistic in inappropriate situations. *Econometrica 34*, 235–238. [155]

Nickell, S. J. (1990). Inflation and the UK labour market. *Oxford Review of Economic Policy 6*, 26–35. [104]

Nielsen, H. B. (2009). Comment on 'the long-run determinants of UK wages, 1860–2004'. *Journal of Macroeconomics 31*, 29–34. [112, 114]

Pember Reeves, M. (1913). *Round about a Pound a Week*. London: Republished in 2008 by Persephone Books. [26, 104]

Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica 25*, 283–299. [53, 104, 163]

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B 31*, 350–371. [79]

Reis, R. (2006). Inattentive producers. *Review of Economic Studies 73*, 793–821. [108]

Rowlatt, P. A. (1988). Analysis of the recent path of UK inflation. *Oxford Bulletin of Economics and Statistics 50*, 335–360. [104]

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In P. E. Hart, G. Mills, and J. K. Whitaker (Eds.), *Econometric Analysis for National Economic Planning*, Volume 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co. [104, 108, 150]

Sargan, J. D. (1980). A model of wage-price inflation. *Review of Economic Studies 47*, 979–1012. [104]

Shadman-Mehta, F. (1995). An empirical study of the determinants of real wages and employment: The Phillips curve revisited. Unpublished thesis, Université Catholique de Louvain, Belgium. [163]

Sleeman, A. (1981). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1851-1979. Paper presented to the Atlantic Economic Society, LSE, London. [163]

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*, 65–94. [11]

Starr, R. M. (2012). *Why Is There Money? Walrasian General Equilibrium Foundations of Monetary Theory*. Aldershot: Edward Elgar. [126]

Student (1908). On the probable error of the mean. *Biometrika 6*, 1–25. [64]

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*, 208–218. [39]

Thomas, J. J. (1984). Wages and prices in the United Kingdom, 1862-1913: A re-examination of the Phillips curve. Presentation, ESRC Quantitative Economic History Study Group, Oxford. [163]

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–838. [79]

Author Index

Index