The model could, of course, be used to perform other forecasting experiments. For example, the effects of a change in the propensity to consume could be measured by changing the coefficient, multiplying disposable income net of transfer payments in the consumption equation, and then simulating the model. Readers might want to use this model to perform their own simulation experiments.

EXERCISES

13.1 Show that if $A_1\lambda_1^t$ and $A_2\lambda_2^t$ are both transient solutions to a model, i.e., both satisfy an equation such as (13.5), then the sum $A_1\lambda_1^t + A_2\lambda_2^t$ must also be a solution.

13.2 Consider the following simple multiplier-accelerator macroeconomic model:

$$C_t = a_1 + a_2 Y_{t-1} \qquad I_t = b_1 + b_2(C_t - C_{t-1}) \qquad Y_t = C_t + I_t + G_t$$

Note that investment is now a function of changes in consumption, rather than of changes in total GNP.

(a) Determine the characteristic equation for this model, and find the associated characteristic roots.

(b) Find the relationships between values of a_2 and b_2 that determine what kind of solution the model will have. Draw a diagram that corresponds to that of Fig. 13.1.

(c) What is the impact multiplier corresponding to a change in G_t? What is the *total* long-run multiplier corresponding to a change in G_t?

13.3 The following equations describe a simple "cobweb" model of a competitive market:

Demand: $\qquad Q_t^D = a_1 + a_2 P_t \qquad a_2 < 0$

Supply: $\qquad Q_t^S = b_1 + b_2 P_{t-1} \qquad b_2 > 0$

When the market is in equilibrium, $Q_t^D = Q_t^S$. Now suppose that the market is temporarily out of equilibrium, i.e., that $Q_t^D \neq Q_t^S$ temporarily.

(a) Show that the price will converge stably to an equilibrium value if $b_2/a_2 < 1$.

(b) Show that the path to equilibrium will be oscillatory if $b_2 > 0$ and will not be oscillatory if $b_2 < 0$.

TIME-SERIES MODELS

In the first two parts of this book we saw how econometric models—both single-equation regression models and multi-equation models—can be constructed and used to explain and forecast the future movements of one or more variables. In Part Three we are again interested in constructing models and using them for forecasting, but these models are quite different from those that we worked with earlier. We no longer predict future movements in a variable by relating it to a set of other variables in a causal framework; instead we base our prediction solely on the past behavior of the variable.

As an example, consider the time series $y(t)$ drawn in the figure on page 414, which might represent the historical performance of some economic or business variable—a stock market index, an interest rate, a production index, or perhaps the daily sales volume for some commodity. $y(t)$ might have moved up or down partly in response to changes in prices, personal income, and interest rates (or so we might believe). However, much of its movement may have been due to factors that we cannot explain, such as the weather, changes in taste, or simply seasonal (or aseasonal) cycles in spending.

It may be difficult or impossible to explain the movement of $y(t)$ through the use of a structural model. This might happen if, for example, data are not available for those explanatory variables which are believed to affect $y(t)$. Or if data were available, the estimation of a regression model for $y(t)$ might result in standard errors that are so large as to make most of the estimated coefficients insignificant and the standard error of forecast unacceptably large.

Even if we could estimate a statistically significant regression equation for $y(t)$, the result may not be useful for forecasting purposes. To obtain a forecast for $y(t)$ from a regression equation, explanatory variables that are not lagged

413
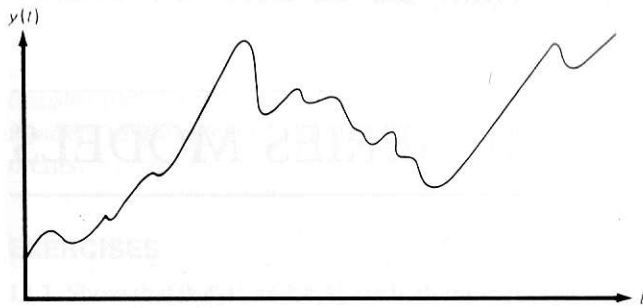
must themselves be forecasted, and this may be more difficult than forecasting $y(t)$ itself. The standard error of forecast for $y(t)$ with future values of the explanatory variables known may be small. However, when the future values of the explanatory variables are unknown, their forecast errors may be so large as to make the total forecast error for $y(t)$ too large to be acceptable.

Thus there are situations where we seek an alternative means of obtaining a forecast of $y(t)$. Can we observe the time series in the figure and draw some conclusions about its *past* behavior that would allow us to infer something about its probable *future* behavior? For example, is there some kind of overall upward trend in $y(t)$ which, because it has dominated the past behavior of the series, might dominate its future behavior? Or does the series exhibit cyclical behavior which we could extrapolate into the future? If systematic behavior of this type is present, we can attempt to construct a *model* for the time series which does not offer a structural explanation for its behavior in terms of other variables but does replicate its past behavior in a way that might help us forecast its future behavior. A *time-series model* accounts for patterns in the past movements of a variable and uses that information to predict its future movements. In a sense a time-series model is just a sophisticated method of extrapolation. Yet, as we will see in this part of the book, it sometimes provides an effective tool for forecasting.

In this book we have divided forecasting models into three general classes, each of which involves a different level of comprehension about the real world processes that one is trying to model. In Part One we discussed single-equation regression models, where the variable of interest is explained by a single function (linear or nonlinear) of explanatory variables. In Part Two we examined multi-equation models, where two or more endogenous variables are related to each other (and perhaps to one or more exogenous variables) through a set of equations, which can be solved simultaneously to produce forecasts over time. In this third part of the book we focus on time-series models, in which we have no structural knowledge about the real world causal relationships that affect the variable we are trying to forecast.

Often a choice must be made as to which type of model should be developed to best make a forecast. This choice may be difficult and will depend not only on how much we know about the workings of the real world process but also on how much time and energy we can afford to spend on the modeling process. The development of a single-equation regression model may not be too difficult, but the construction of a multi-equation simulation model might require large expenditures of time and effort. The gains that result might include a better understanding of the relationships and structure involved as well as the ability to make a better forecast. However, in some cases these gains may be outweighed by the cost involved.

A time-series model will usually be chosen when little is known about the determinants of the variable of primary concern and a sufficiently large amount of data is available to construct a time series of reasonable length. An example might be to forecast on a weekly or monthly basis a cyclical series for the production of a commodity. One choice would be to build a regression model, a second choice would be to build a time-series model, and a third choice is to *combine* time-series analysis with regression analysis. As we will see, one can construct a regression model in which, say, an interest rate is related to several economic variables and then construct a time-series model to explain the behavior of the residual term from the regression.

The following chapters will present an introduction to the science and art of developing time-series models for purposes of forecasting. The models that we deal with are only a subset of a broader class of models and techniques for data analysis that falls under the general rubric of time-series analysis. We do not discuss, for example, recent developments in the theory of *spectral analysis* and its application to economic modeling and forecasting. Instead we concentrate on a class of linear time-series models introduced by Box and Jenkins that have found wide application to economic and business forecasting.[1]

Since time-series analysis builds on the development of the single-equation regression model, we treat time-series models in the last part of the book, even though they are the "simplest" class of models in terms of their explanation of the real world. To forecast a short-term interest rate, we might use a regression model to relate that variable to GNP, prices, and the money supply. A time series for interest rates would relate that variable to its past values and to variables that describe the random nature of its past behavior. The model, like most regression models, is an equation containing a set of coefficients that must be estimated. Unlike regression models, however, the equation is usually nonlinear in the coefficients, so that a nonlinear version of ordinary least squares is necessary for estimation purposes.

Part Three begins with a brief survey of simple extrapolation methods (in effect deterministic models of time series), as well as methods for smoothing and seasonally adjusting time series. Extrapolation techniques have been used widely for many years and for some applications provide a simple and yet adequate means of forecasting. Smoothing and seasonal adjustment are also useful techniques, which in many instances can facilitate the forecasting or interpretation of a time series.

---

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (San Francisco: Holden-Day, 1970).

In Chapter 15 we present a brief introduction to the nature of stochastic time series. We discuss how stochastic processes are generated, what they look like, and most important, how they are described. We also discuss some of the characteristics of stochastic processes and in particular develop the concept of stationarity. Then we describe autocorrelation functions and show how they can be used as a means of describing time series and as a tool for testing their properties. Finally, we discuss methods of testing for stationarity, and we discuss the concept of *co-integrated* time series. The concepts and tools developed in this chapter are essential to the discussion of time-series models in the chapters that follow.

Chapter 16 develops linear models for time series, including moving average models, autoregressive models, and mixed autoregressive–moving average models for stationary time series. We show how some nonstationary time series can be differenced one or more times so as to produce a stationary series. This enables us to develop a general integrated autoregressive–moving average model (ARIMA model). Finally, we show how autocorrelation functions can be used to specify and characterize a time-series model.

Chapters 17 and 18 deal with use of time-series models to make forecasts. Chapter 17 explains how parameters of a time-series model are estimated and how a specification of the model can be verified. Chapter 18 discusses how the model can then be used to produce a forecast. We also show how time series are adaptive in nature, i.e., how they produce forecasts in a way that adapts to new information. The last part of Chapter 18 deals with forecast errors and shows how confidence intervals can be determined for forecasts.

The last chapter of Part Three develops some examples of applications of time-series models to economic and business forecasting. Here we lead the reader step by step through the construction of several time-series models and their application to forecasting problems.

SMOOTHING AND EXTRAPOLATION OF TIME SERIES

As explained in the introduction to Part Three, a time-series model is a sophisticated method of extrapolating data. There are times, however, when less sophisticated methods of extrapolation can be used for forecasting purposes. For example, projections for a large number of time series might be needed quickly, so that time and resources do not permit the use of formal modeling techniques, or there might be reason to believe that a particular time series follows a simple trend, thus obviating the need for a more complicated model. We therefore begin by discussing some simple (and not so simple) methods of extrapolation. These extrapolation techniques represent *deterministic models* of time series.

There are also situations when it is desirable to *smooth* a time series and thereby eliminate some of the more volatile short-term fluctuations. Smoothing might be done before making a forecast or simply to make the time series easier to analyze and interpret. Smoothing might also be done to remove seasonal fluctuations, i.e., to *deseasonalize* (or *seasonally adjust*) a time series. We will discuss smoothing and seasonal adjustment in the second section of this chapter.

14.1 SIMPLE EXTRAPOLATION MODELS

We begin with simple models that can be used to forecast a time series on the basis of its past behavior. These models are *deterministic* in that no reference is made to the sources or nature of the underlying randomness in the series. Essentially the models involve extrapolation techniques that have been standard tools of the trade in economic and business forecasting for years. Although they usually do not provide as much forecasting accuracy as the modern stochastic
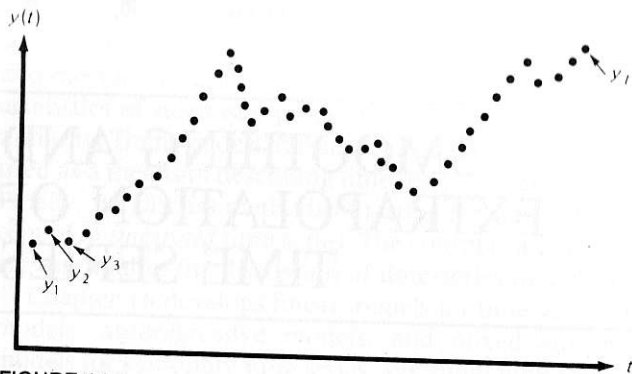
FIGURE 14.1
Discrete time series.

time-series models, they often provide a simple, inexpensive, and still quite acceptable means of forecasting.

Most of the series that we encounter are not continuous in time; instead they consist of discrete observations made at regular intervals of time. A typical time series might be given by Fig. 14.1. We denote the values of that series by y_t, so that y_1 represents the first observation, y_2 the second, and y_T the *last* observation for the series.[1] Our objective is to model the series y_t and use that model to forecast y_t beyond the last observation y_T. We denote the forecast one period ahead by $\hat{y}_{T+1}$, two periods ahead by $\hat{y}_{T+2}$, and l periods ahead by $\hat{y}_{T+l}$.

If the number of observations is not too large, the simplest and most complete representation of y_t would be given by a polynomial whose degree is 1 less than the number of observations; i.e., we could describe y_t by a continuous function of time $f(t)$, where

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \tag{14.1}$$

and $n = T - 1$. Such a polynomial (if the a's are chosen correctly) will pass through *every point* in the time series y_t. Thus, we can be sure that $f(t)$ will equal y_t at every time t from 1 to T. Can we, however, have any confidence that a *forecast* of y_t generated by $f(t)$ will be at all close to its actual future value? For example, will the forecast

$$f(T + 1) = a_0 + a_1(T + 1) + a_2(T + 1)^2 + \cdots + a_{T-1}(T + 1)^{T-1} = \hat{y}_{T+1}$$

be close to the actual future value y_{T+1}? Unfortunately, we have no way of answering this question without additional prior information. The difficulty with the model given by Eq. (14.1) is that it does not describe y_t; it merely *reproduces* y_t. It does not capture any characteristics of y_t that might repeat themselves in the future. Thus, although $f(t)$ correlates perfectly with y_t, it is of little use for forecasting.

[1] In Part Three of the book we use small letters, for example, y_t, to denote time series.

14.1.1 Simple Extrapolation Models

One basic characteristic of y_t is its long-run growth pattern. If we believe that this upward trend exists and will continue (and there may not be any reason why we should), we can construct a simple model that describes that trend and can be used to forecast y_t.

The simplest extrapolation model is the *linear trend model*. If we believe that a series y_t will increase in constant absolute amounts each time period, we can predict y_t by fitting the trend line

$$y_t = c_1 + c_2 t \tag{14.2}$$

where t is time and y_t is the value of y at time t. t is usually chosen to equal 0 in the base period (first observation) and to increase by 1 during each successive period. For example, if we determine by regression that

$$y_t = 27.5 + 3.2t \tag{14.3}$$

we can predict that the value of y in period $t + 1$ will be 3.2 units higher than the previous value.

It may be more realistic to assume that the series y_t grows with constant percentage increases, rather than constant absolute increases. This assumption implies that y_t follows an *exponential growth curve*:

$$y_t = f(t) = A e^{rt} \tag{14.4}$$

Here A and r would be chosen to maximize the correlation between $f(t)$ and y_t. A forecast one period ahead would then be given by

$$\hat{y}_{T+1} = A e^{r(T+1)} \tag{14.5}$$

and l periods ahead by

$$\hat{y}_{T+l} = A e^{r(T+l)} \tag{14.6}$$

This is illustrated in Fig. 14.2. The parameters A and r can be estimated by taking the logarithms of both sides of Eq. (14.4) and fitting the log-linear regression equation[2]

$$\log y_t = c_1 + c_2 t \tag{14.7}$$

where $c_1 = \log A$ and $c_2 = r$.

[2] Note that in the exponential growth model the logarithm of y_t is assumed to grow at a constant rate. If $y_{t+1} = A e^{rt}$, then $y_{t+1}/y_t = e^r$, and $\log y_{t+1} - \log y_t = r$.
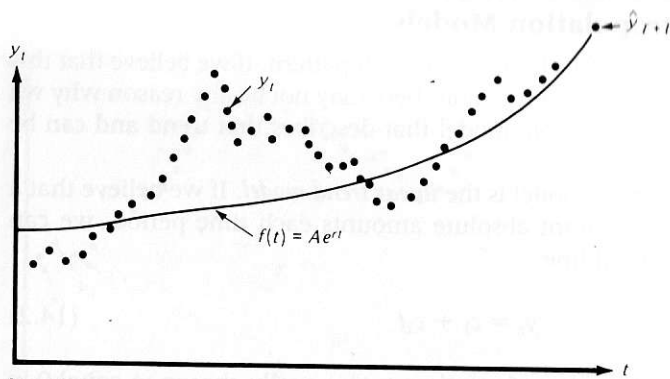
FIGURE 14.2
Exponential growth curve.

A third extrapolation method is based on the *autoregressive trend model*

$$y_t = c_1 + c_2 y_{t-1} \tag{14.8}$$

In using such an extrapolation procedure, one has the option of fixing $c_1 = 0$, in which case c_2 represents the rate of change of the series y. If, on the other hand, c_2 is set equal to 1, with c_1 not equal to 0, the extrapolated series will increase by the same absolute amount each time period. The autoregressive trend model is illustrated in Fig. 14.3 for three different values of c_2 (in all cases $c_1 = 1$).

A variation of this model is the *logarithmic autoregressive trend model*

$$\log y_t = c_1 + c_2 \log y_{t-1} \tag{14.9}$$

FIGURE 14.3
Autoregressive trend model.
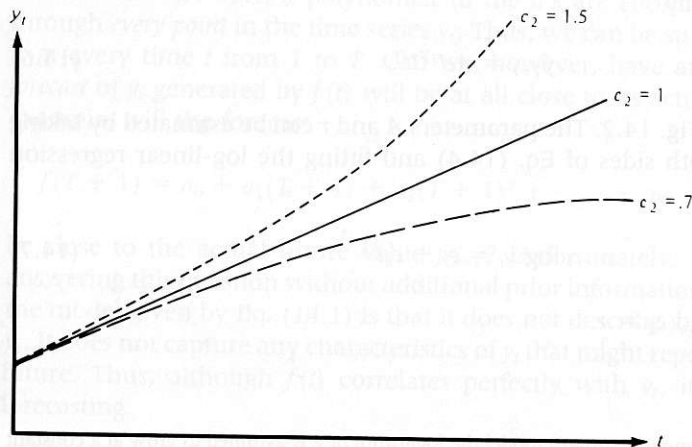
If c_1 is fixed to be 0, then the value of c_2 is the compounded rate of growth of the series y. Both linear and compound extrapolation based on the autoregressive model are commonly used as a simple means of forecasting.

Note that the four models described above basically involve regressing y_t (or $\log y_t$) against a function of time (linear or exponential) and/or itself lagged. Alternative models can be developed by making the function slightly more complicated. As examples, let us examine two other simple extrapolation models, the *quadratic trend model* and the *logistic growth curve*.

The quadratic trend model is a simple extension of the linear trend model and involves adding a term in t^2

$$y_t = c_1 + c_2 t + c_3 t^2 \tag{14.10}$$

If c_2 and c_3 are both positive, y_t will always be increasing, but even more rapidly as time goes on. If c_2 is negative and c_3 positive, y_t will at first decrease but later increase. If both c_2 and c_3 are negative, y_t will always decrease. The various cases are illustrated in Fig. 14.4 ($c_1 > 0$ in each case). Note that even if the data show that y_t has generally been increasing over time, estimation of Eq. (14.10) might yield a positive value for c_3 but a negative value for c_2. This can occur (as shown in Fig. 14.4) because the data usually only span a portion of the trend curve.

A somewhat more complicated model, at least in terms of its estimation, is the *logistic curve*, given by

$$y_t = \frac{1}{k + ab^t} \qquad b > 0 \tag{14.11}$$

FIGURE 14.4
Quadratic trend model.
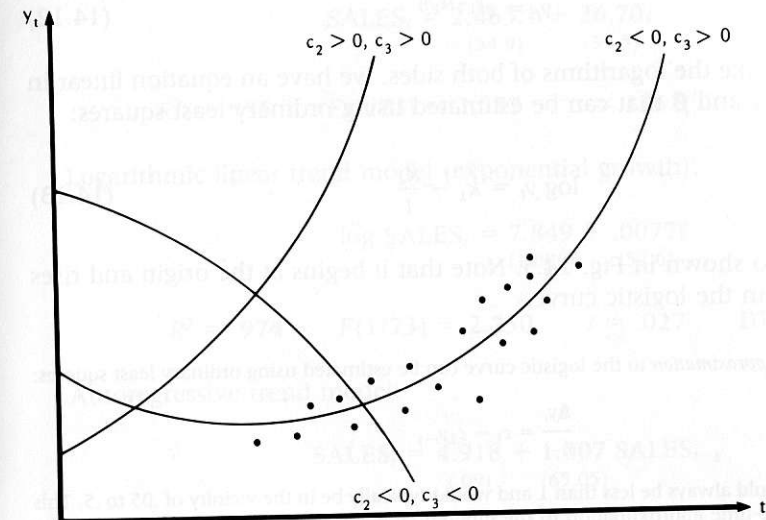
FIGURE 14.5
S-shaped curves.

This equation is nonlinear in the parameters (k, a, and b) and therefore must be estimated using a nonlinear estimation proced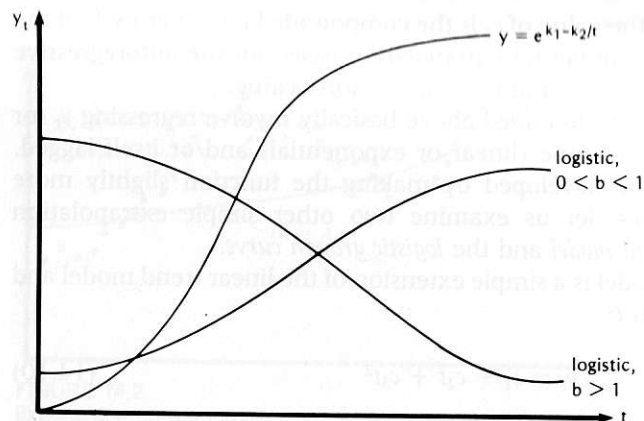ure. While this can add computational expense, there are some cases in which it is worth it. As shown in Fig. 14.5, Eq. (14.11) represents an S-shaped curve which might be used to represent the sales of a product that will someday saturate the market (so that the total stock of the good in circulation will approach some plateau, or, equivalently, additional sales will approach zero).[3]

Other S-shaped curves can be used in addition to the logistic curve. One very simple function with an S shape that can be used to model sales saturation patterns is given by

$$y_t = e^{k_1 - (k_2/t)} \qquad (14.12)$$

Note that if we take the logarithms of both sides, we have an equation linear in the parameters α and β that can be estimated using ordinary least squares:

$$\log y_t = k_1 - \frac{k_2}{t} \qquad (14.13)$$

This curve is also shown in Fig. 14.5. Note that it begins at the origin and rises more steeply than the logistic curve.

[3] The following *approximation* to the logistic curve can be estimated using ordinary least squares:

$$\frac{\Delta y_t}{y_{t-1}} = c_1 - c_2 y_{t-1}$$

The parameter c_2 should always be less than 1 and would typically be in the vicinity of .05 to .5. This equation is a discrete-time approximation to the differential equation $dy/dt = c_2 y(c_1 - y)$, and the *solution* to this differential equation has the form of Eq. (14.11).

Example 14.1 Forecasting Department Store Sales In this example simple extrapolation models are used to forecast monthly retail sales of department stores. The time series is listed below, where monthly observations are seasonally adjusted and cover the period from January 1968 to March 1974, the units of measurement are millions of dollars, and the source of the data is the U.S. Department of Commerce.

	1968	1969	1970	1971	1972	1973	1974
January	2,582	2,839	3,034	3,287	3,578	4,121	4,456
February	2,621	2,876	3,029	3,342	3,650	4,233	4,436
March	2,690	2,881	3,045	3,336	3,664	4,439	4,699
April	2,635	2,967	3,066	3,427	3,643	4,167	
May	2,676	2,944	3,077	3,413	3,838	4,326	
June	2,714	2,939	3,046	3,503	3,792	4,329	
July	2,834	3,014	3,094	3,472	3,899	4,423	
August	2,789	3,031	3,053	3,511	3,845	4,351	
September	2,768	2,995	3,071	3,618	4,007	4,406	
October	2,785	2,998	3,186	3,554	4,092	4,357	
November	2,886	3,012	3,167	3,641	3,937	4,485	
December	2,842	3,031	3,230	3,607	4,008	4,445	

One might wish to forecast monthly sales for April, May, and the months following in 1974. For this example, we extrapolate sales for April 1974. The results of four regressions associated with four of the trend models described above are listed below. Standard regression statistics are shown with t statistics in parentheses:

Linear trend model:

$$\text{SALES}_t = 2,463.1 + 26.70t \qquad (14.14)$$
$$\quad\ (84.9) \qquad (39.5)$$

$$R^2 = .955 \qquad F(1/73) = 1,557 \qquad s = 126.9 \qquad DW = .38$$

Logarithmic linear trend model (exponential growth):

$$\log \text{SALES}_t = 7.849 + .0077t \qquad (14.15)$$
$$\qquad\quad (1,000) \qquad (52.6)$$

$$R^2 = .974 \qquad F(1/73) = 2,750 \qquad s = .027 \qquad DW = .56$$

Autoregressive trend model:

$$\text{SALES}_t = 4.918 + 1.007\ \text{SALES}_{t-1} \qquad (14.16)$$
$$\qquad (.09) \qquad (65.05)$$

$$R^2 = .983 \qquad F(1/72) = 3,829 \qquad s = 78.07 \qquad DW = 2.82$$

Logarithmic autoregressive trend model:

$$\log SALES_t = .0188 + .9987 \log SALES_{t-1} \qquad (14.17)$$
$$ (.16) \qquad (70.37)$$

$$R^2 = .985 \qquad F(1/72) = 4,524 \qquad s = .021 \qquad DW = 2.80$$

In the first regression, a time variable running from 0 to 74 was constructed and then used as the independent variable. When $t = 75$ is placed in the right-hand side of the equation

$$SALES = 2,463.1 + 26.70t \qquad (14.18)$$

the resulting forecast is 4,465.8. The use of the second log-linear equation yields a forecast of 4,551.5. The third regression, based on an autoregressive

FIGURE 14.6
Simulated and actual sales.

(a)

(b)
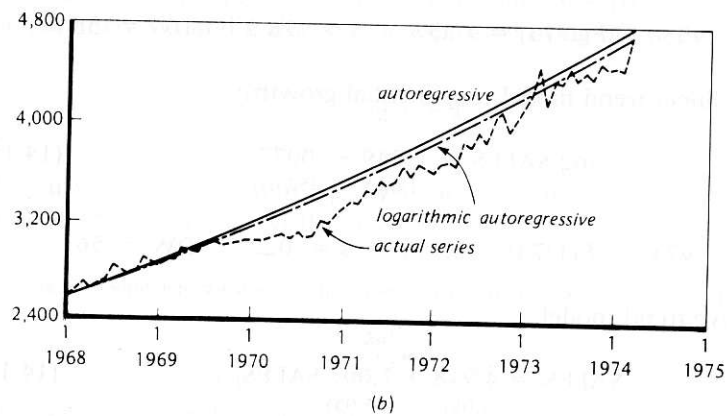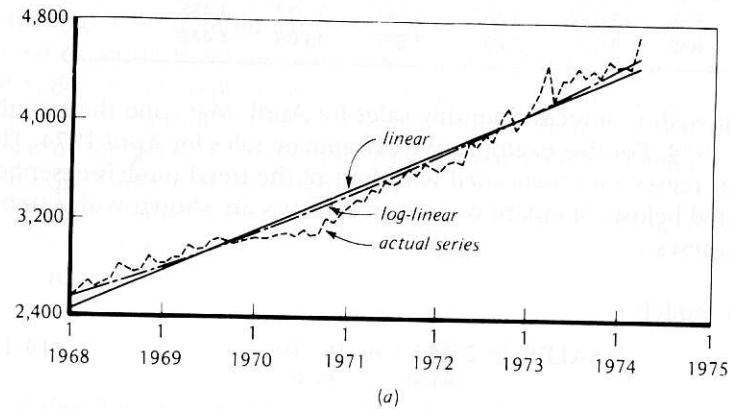
process, yields an extrapolated value for April 1974 of 4,736.8:

$$4,736.8 \approx 4.92 + 1.007 \times 4,699$$

If the constant term were dropped from Eq. (14.17), the extrapolated value would be 4,738.24. The fourth regression result is based on the logarithmic autoregressive model. The extrapolated value in this case is 4,735.6. If one were to calculate a compounded growth rate for the series and to extrapolate on the basis that the growth rate remains unchanged, the extrapolated value would be 4,739.3.

The simulated and actual series are plotted for each of the four extrapolation models in Fig. 14.6a and b. One can see from the figure that the two autoregressive models are closer to the actual series at the end of the period. Of course, other trend models could be used to extrapolate the data. For example, the reader might try to calculate a forecast based on a quadratic trend model (see Exercise 14.1).

Simple extrapolation methods such as those used in the preceding example are frequently the basis for making casual long-range forecasts of variables ranging from GNP to population to pollution indices. Although they can be useful as a way of quickly formulating initial forecasts, they usually provide little forecasting accuracy. The analyst who estimates an extrapolation model is at least advised to calculate a standard error of forecast and forecast confidence interval following the methods presented in Chapter 8. More important, one should realize that there are alternative models that can be used to obtain forecasts with smaller standard errors.

14.1.2 Moving Average Models

Another class of deterministic models that are often used for forecasting consists of *moving average models*. As a simple example, assume that we are forecasting a monthly time series. We might use the model

$$f(t) = \tfrac{1}{12}(y_{t-1} + y_{t-2} + \cdots + y_{t-12}) \qquad (14.19)$$

Then, a forecast one period ahead would be given by

$$\hat{y}_{T+1} = \tfrac{1}{12}(y_T + y_{T-1} + \cdots + y_{T-11}) \qquad (14.20)$$

The moving average model is useful if we believe that a likely value for our series next month is a simple average of its values over the past 12 months. It may be unrealistic, however, to assume that a good forecast of y_t would be given

by a simple average of its past values. It is often more reasonable to have more recent values of y_t play a greater role than earlier values. In such a case recent values should be weighted more heavily in the moving average. A simple model that accomplishes this is the *exponentially weighted moving average* (EWMA) model:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \cdots$$

$$= \alpha \sum_{\tau=0}^{\infty} (1 - \alpha)^\tau y_{T-\tau} \qquad (14.21)$$

Here α is a number between 0 and 1 that indicates how heavily we weight recent values relative to older ones. With $\alpha = 1$, for example, our forecast becomes

$$\hat{y}_{T+1} = y_T \qquad (14.22)$$

and we ignore any values of y that occurred before y_T. As α becomes smaller, we place greater emphasis on more distant values of y. Note that Eq. (14.21) represents a true average, since

$$\alpha \sum_{\tau=0}^{\infty} (1 - \alpha)^\tau = \frac{\alpha}{1 - (1 - \alpha)} = 1 \qquad (14.23)$$

so that the weights indeed sum to unity.

The reader might suspect that if the series has an upward (downward) trend, the EWMA model will underpredict (overpredict) future values of y_t. This will indeed be the case, since the model averages past values of y_t to produce a forecast. If y_t has been growing steadily in the past, the EWMA forecast $\hat{y}_{T+1}$ will thus be *smaller* than the most recent value y_T, and if the series continues to grow steadily in the future, $\hat{y}_{T+1}$ will be an underprediction of the true value y_{T+1}. Thus one ought to remove any trend from the data before using the EWMA technique. Once an untrended initial forecast has been made, the trend term can be added to obtain a final forecast.

If we want to make a forecast $\hat{y}_{T+1}$ more than one period ahead using an exponentially weighted moving average model, we can modify Eq. (14.21) to include a weighted average of the more recent short-run forecasts $\hat{y}_{T+l-1}, \hat{y}_{T+l-2}, \ldots, \hat{y}_{T+1}$. This logical extension of the EWMA model is given by

$$\hat{y}_{T+l} = \alpha\hat{y}_{T+l-1} + \alpha(1 - \alpha)\hat{y}_{T+l-2} + \cdots + \alpha(1 - \alpha)^{l-2}\hat{y}_{T+1}$$

$$+ \alpha(1 - \alpha)^{l-1}y_T + \alpha(1 - \alpha)^l y_{T-1} + \alpha(1 - \alpha)^{l+1}y_{T-2}$$

$$+ \alpha(1 - \alpha)^{l+2}y_{T-3} + \cdots \qquad (14.24)$$

As an example, consider a forecast two periods ahead ($l = 2$), which would be given by

$$\hat{y}_{T+2} = \alpha\hat{y}_{T+1} + \alpha(1 - \alpha)y_T + \alpha(1 - \alpha)^2 y_{T-1} + \cdots$$

$$= \alpha[\alpha y_T + \alpha(1 - \alpha)y_{T-1} + \cdots] + \alpha(1 - \alpha)y_T$$

$$+ \alpha(1 - \alpha)^2 y_{T-1} + \cdots$$

$$= \alpha^2 \sum_{\tau=0}^{\infty} (1 - \alpha)^\tau y_{T-\tau} + \alpha(1 - \alpha) \sum_{\tau=0}^{\infty} (1 - \alpha)^\tau y_{T-\tau}$$

$$= \alpha \sum_{\tau=0}^{\infty} (1 - \alpha)^\tau y_{T-\tau} \qquad (14.25)$$

Note that the two-period forecast is the same as the one-period forecast. The weightings on $y_T, y_{T-1}, \ldots$, in the EWMA model are the same as they were before, but we are now extrapolating the average ahead an extra period. In fact, it is not difficult to show (see Exercise 14.4) that the l-period forecast $\hat{y}_{T+l}$ is also given by Eq. (14.25).

The moving average forecasts represented by Eqs. (14.20), (14.21), and (14.24) are all *adaptive forecasts*. By "adaptive" we mean that they automatically adjust themselves to the most recently available data. Consider, for example, a simple four-period moving average. Suppose y_{20} in Fig. 14.7 represents the most recent data point. Then our forecast will be given by

$$\hat{y}_{21} = \tfrac{1}{4}(y_{20} + y_{19} + y_{18} + y_{17}) \qquad (14.26)$$

and a forecast two periods ahead will be given by

$$\hat{y}_{22} = \tfrac{1}{4}(\hat{y}_{21} + y_{20} + y_{19} + y_{18}) = \tfrac{5}{16}y_{20} + \tfrac{5}{16}y_{19} + \tfrac{5}{16}y_{18} + \tfrac{1}{16}y_{17} \qquad (14.27)$$

FIGURE 14.7
Adaptive forecasts.
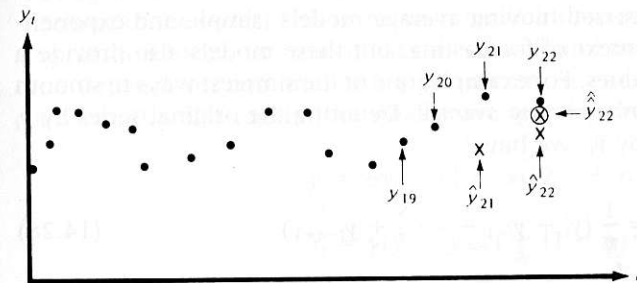
These forecasts are represented by crosses in Fig. 14.7. If y_{21} were known, we would forecast y_{22} one period ahead as

$$\hat{y}_{22} = \tfrac{1}{4}(y_{21} + y_{20} + y_{19} + y_{18})$$

This forecast is represented by a circled cross in Fig. 14.7. Now suppose that the *actual* value of y_{21} turns out to be larger than the predicted value, i.e.,

$$y_{21} > \hat{y}_{21}$$

The actual value of y_{22} is, of course, not known, but we would expect that $\hat{\hat{y}}_{22}$ would provide a better forecast than $\hat{y}_{22}$ because of the extra information used in the adaptive process. The EWMA forecast would exhibit the same adaptive behavior.

Although the moving average models described above are certainly useful, they do not provide us with information about *forecast confidence*. The reason is that no regression is used to estimate the model, so that we cannot calculate standard errors, nor can we describe or explain the stochastic (or unexplained) component of the time series. It is this stochastic component that creates the error in our forecast. Unless the stochastic component is explained through the modeling process, little can be said about the kinds of forecast errors that might be expected.

14.2 SMOOTHING AND SEASONAL ADJUSTMENT

Smoothing techniques provide a means of removing or at least reducing volatile short-term fluctuations in a time series. This can be useful since it is often easier to discern trends and cyclical patterns and otherwise visually analyze a smoothed series. Seasonal adjustment is a special form of smoothing; it removes seasonal (cyclical) oscillations from the series rather than irregular short-term fluctuations.

14.2.1 Smoothing Techniques

In the last section we discussed moving average models (simple and exponentially weighted) in the context of forecasting, but these models also provide a basis for smoothing time series. For example, one of the simplest ways to smooth a series is to take an *n-period moving average*. Denoting the original series by y_t and the smoothed series by $\bar{y}_t$, we have

$$\bar{y}_t = \frac{1}{n}(y_t + y_{t-1} + \cdots + y_{t-n+1}) \tag{14.28}$$

Of course, the larger the n the smoother the $\bar{y}_t$ will be. One problem with this moving average is that it uses only *past* (and current) values of y_t to obtain each

value of $\bar{y}_t$. This problem is easily remedied by using a *centered moving average*. For example, a five-period centered moving average is given by

$$\bar{y}_t = \tfrac{1}{5}(y_{t+2} + y_{t+1} + y_t + y_{t-1} + y_{t-2}) \tag{14.29}$$

Exponential smoothing simply involves the use of the exponentially weighted moving average model for smoothing. (Recall that this model assigns heavier weights to recent values of y_t.) The exponentially smoothed series $\bar{y}_t$ is given by

$$\bar{y}_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \tag{14.30}$$

where the summation in Eq. (14.30) extends all the way back through the length of the series. In fact, $\bar{y}_t$ can be calculated much more easily if we write

$$(1-\alpha)\bar{y}_{t-1} = \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \tag{14.31}$$

Now subtracting Eq. (14.31) from Eq. (14.30) we obtain a recursive formula for the computation of $\bar{y}_t$:

$$\bar{y}_t = \alpha y_t + (1-\alpha)\bar{y}_{t-1} \tag{14.32}$$

Note that the closer α is to 1 the more heavily the current value of y_t is weighted in generating $\bar{y}_t$. Thus smaller values of α imply a more heavily smoothed series.

Sometimes one might wish to heavily smooth a series but not give very much weight to past data points. In such a case the use of Eq. (14.32) with a small value of α (say .1) would not be acceptable. Instead one can apply *double exponential smoothing*. As the name implies, the singly smoothed series $\bar{y}_t$ from Eq. (14.32) is just smoothed again:

$$\bar{\bar{y}}_t = \alpha \bar{y}_t + (1-\alpha)\bar{\bar{y}}_{t-1} \tag{14.33}$$

In this way a larger value of α can be used, and the resulting series $\bar{\bar{y}}_t$ will still be heavily smoothed.

The simple exponential smoothing formula of Eq. (14.32) can also be modified by incorporating average *changes* in the long-run trend (secular increase or decline) of the series. This is the basis for *Holt's two-parameter exponential smoothing* method.[4] Now the smoothed series $\bar{y}_t$ is found from two recursive equations and depends on two smoothing parameters, α and γ, both of which must lie between 0 and 1 (again, the smaller are α and γ the heavier the smoothing):

$$\bar{y}_t = \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + r_{t-1}) \tag{14.34}$$

$$r_t = \gamma(\bar{y}_t - \bar{y}_{t-1}) + (1-\gamma)r_{t-1} \tag{14.35}$$

[4] C. C. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages," unpublished research report, Carnegie Institute of Technology, Pittsburgh, 1957.

Here r_t is a smoothed series representing the trend, i.e., average rate of increase, in the smoothed series $\tilde{y}_t$. This trend is added in when computing the smoothed series $\tilde{y}_t$ in Eq. (14.34), thereby preventing $\tilde{y}_t$ from deviating considerably from recent values of the original series y_t. This is particularly useful if the smoothing method is going to be used as a basis for forecasting. An l-period forecast can be generated from Eqs. (14.34) and (14.35) using

$$\hat{y}_{T+l} = \tilde{y}_T + lr_T \tag{14.36}$$

Thus the l-period forecast takes the most recent smoothed value $\tilde{y}_T$ and adds in an expected increase lr_T based on the (smoothed) long-run trend. (If the data have been detrended, the trend should be added back to the forecast.)

Smoothing methods tend to be *ad hoc*, particularly when they are used to generate forecasts. One problem is that we have no way of determining the "correct" values of the smoothing parameters, so that their choice becomes somewhat arbitrary. If our objective is simply to smooth the series to make it easier to interpret or analyze, then this is not really a problem, since we can choose the smoothing parameters to give us the extent of smoothing desired. We must be careful, however, when using an equation like Eq. (14.36) for forecasting and recognize that the resulting forecast will be somewhat arbitrary.[5]

Example 14.2 Monthly Housing Starts The time series for monthly housing starts in the United States provides a good example for the application of smoothing and seasonal adjustment methods.[6] The series fluctuates considerably and also exhibits strong seasonal variation. In this example we smooth the series using the moving average and exponential smoothing methods.

We begin by using three- and seven-period centered moving averages to smooth the series; i.e., we generate the smoothed series $\tilde{y}_t$ from the original series y_t using

$$\tilde{y}_t = \frac{1}{n} \sum_{i=0}^{n-1} y_{t+(1/2)(n-1)-i} \tag{14.37}$$

where $n = 3$ or 7. Note that since the moving average is centered, there is no need to detrend the series before smoothing it. The original series, together with the two smoothed series, is shown in Fig. 14.8. Observe that the use of

---

[5] For a detailed treatment of some other smoothing techniques, see C. W. J. Granger and P. Newbold, *Forecasting Economic Time Series* (New York: Academic, 1986), and S. Makridakis and S. C. Wheelwright, *Forecasting Methods and Applications* (New York: Wiley, 1978).
[6] The original data series is in thousands of units per month and is *not* seasonally adjusted.
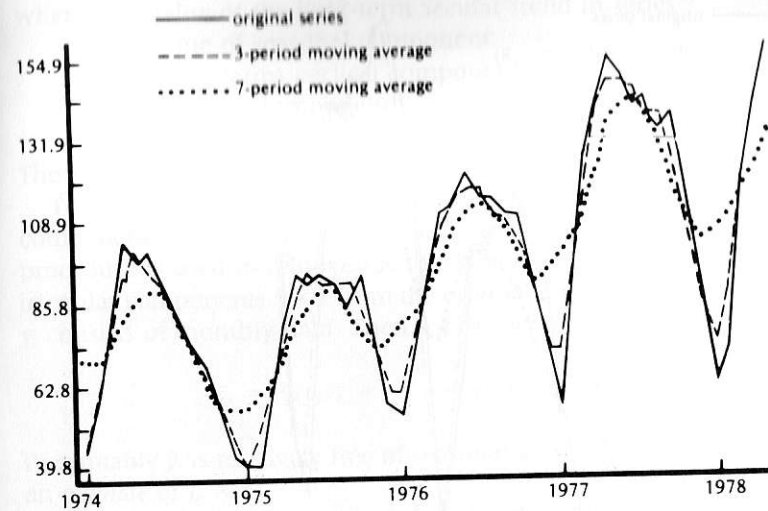
FIGURE 14.8
Smoothing using moving averages.

the seven-period moving average heavily smoothes the series and even eliminates some of the seasonal variation.

We now use the exponential smoothing method, i.e., we apply Eq. (14.32). Since the original series is growing over time and the exponentially weighted moving average is not centered, the smoothed series will underestimate the original series unless we first detrend the series. To detrend the original series we assumed a linear trend (we could of course test alternative time trends), and ran the regression

$$y_t = -156.81 + 1.2083t \qquad R^2 = .360 \tag{14.38}$$
$$\quad\;\;(-3.36) \qquad (5.37)$$

The *residuals* u_t from this regression, that is, $u_t = y_t + 156.81 - 1.2083t$, provide the detrended series.

We next apply exponential smoothing to this detrended series. We use two alternative values of the smoothing parameter, $\alpha = .8$ (light smoothing) and $\alpha = .2$ (heavy smoothing). Finally we take the smoothed detrended series $\tilde{u}_t$ and add the trend back in; i.e., we compute $\tilde{y}_t = \tilde{u}_t - 156.81 + 1.2083t$.

The original series and the smoothed series are shown in Fig. 14.9. Observe from the figure that the seasonal variations, while reduced, are pushed forward by heavy exponential smoothing. This occurs because the exponentially weighted moving average is not centered. Thus if a series shows strong seasonal variations, exponential smoothing should be used only after the series has been seasonally adjusted.

FIGURE 14.9
Smoothing using exponentially weighted moving averages.

where L = value of the long-term secular trend in series
S = value of seasonal component
C = (long-term) cyclical component
I = irregular component

The objective is to eliminate the seasonal component S.

To do this we first try to isolate the combined long-term trend and cyclical components $L \times C$. This cannot be done exactly; instead an *ad hoc* smoothing procedure is used to remove (as much as possible) the combined seasonal and irregular components $S \times I$ from the original series y_t. For example, suppose that y_t consists of monthly data. Then a *12-month average* $\tilde{y}_t$ is computed:

$$\tilde{y}_t = \tfrac{1}{12}(y_{t+6} + \cdots + y_t + y_{t-1} + \cdots + y_{t-5}) \qquad (14.40)$$

Presumably $\tilde{y}_t$ is relatively free of seasonal and irregular fluctuations and is thus an *estimate* of $L \times C$.

We now divide the original data by this estimate of $L \times C$ to obtain an estimate of the combined seasonal and irregular components $S \times I$:

$$\frac{L \times S \times C \times I}{L \times C} = S \times I = \frac{y_t}{\tilde{y}_t} = z_t \qquad (14.41)$$

The next step is to eliminate the irregular component I as completely as possible in order to obtain the seasonal index. To do this, we *average the values of $S \times I$ corresponding to the same month*. In other words, suppose that y_1 (and hence z_1) corresponds to January, y_2 to February, etc., and there are 48 months of data. We thus compute

$$\bar{z}_1 = \tfrac{1}{4}(z_1 + z_{13} + z_{25} + z_{37})$$
$$\bar{z}_2 = \tfrac{1}{4}(z_2 + z_{14} + z_{26} + z_{38}) \qquad (14.42)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\bar{z}_{12} = \tfrac{1}{4}(z_{12} + z_{24} + z_{36} + z_{48})$$

The rationale here is that when the seasonal-irregular percentages z_t are averaged for each month (each quarter if the data are quarterly), the irregular fluctuations will be largely smoothed out.

The 12 averages $\bar{z}_1, \ldots, \bar{z}_{12}$ will then be estimates of the seasonal indices. They should sum close to 12 but will not do so exactly if there is any long-run trend in the data. Final seasonal indices are computed by multiplying the indices in Eq. (14.42) by a factor that brings their sum to 12. (For example, if $\bar{z}_1, \ldots, \bar{z}_{12}$ add to 11.7, multiply each one by 12.0/11.7 so that the revised indices will add to 12.) We denote these final seasonal indices by $\bar{z}_1, \ldots, \bar{z}_{12}$.

The deseasonalization of the original series y_t is now straightforward; just divide each value in the series by its corresponding seasonal index, thereby removing the seasonal component while leaving the other three components.

14.2.2 Seasonal Adjustment

Seasonal adjustment techniques are basically *ad hoc* methods of computing *seasonal indices* (that attempt to measure the seasonal variation in the series) and then using those indices to *deseasonalize* (i.e., seasonally adjust) the series by removing those seasonal variations. National economic data in the United States is usually seasonally adjusted by the *Census II method* (or one of its variants), which was developed by the Bureau of the Census of the U.S. Department of Commerce. The Census II method is a rather detailed and complicated procedure (and is amazingly *ad hoc*), and we therefore will not attempt to describe it here.[7] Instead, we discuss the basic idea that lies behind all seasonal adjustment methods (including Census II) and present a very simple method that in many cases is quite adequate.

Seasonal adjustment techniques are based on the idea that a time series y_t can be represented as the product of four components:

$$y_t = L \times S \times C \times I \qquad (14.39)$$

[7] The Census II method is described in detail in L. Salzman, *Computerized Economic Analysis* (New York: McGraw-Hill, 1968).

Thus the seasonally adjusted series y_t^a is obtained from $y_1^a = y_1/\bar{z}_1$, $y_2^a = y_2/\bar{z}_2$, . . . , $y_{12}^a = y_{12}/\bar{z}_{12}$, $y_{13}^a = y_{13}/\bar{z}_1$, $y_{14}^a = y_{14}/\bar{z}_2$, etc.

Example 14.3 Monthly Housing Starts

Let us now apply the seasonal adjustment technique to our series for monthly housing starts (see Example 14.2). To do this we first compute a 12-month average $\bar{y}_t$ of the original series y_t using Eq. (14.40) and then divide y_t by $\bar{y}_t$, that is, compute $z_t = y_t/\bar{y}_t$. Note that z_t contains (roughly) the seasonal and irregular components of the original series. We remove the irregular component by averaging the values of z_t that correspond to the same month; i.e., we compute $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_{12}$ using Eq. (14.42). We then compute the final *seasonal indices* $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_{12}$ by multiplying the $\bar{z}_1, \ldots, \bar{z}_{12}$ by a factor that brings their sum to 1. The final seasonal indices are as follows:

FIGURE 14.10
Housing starts: seasonal indices.
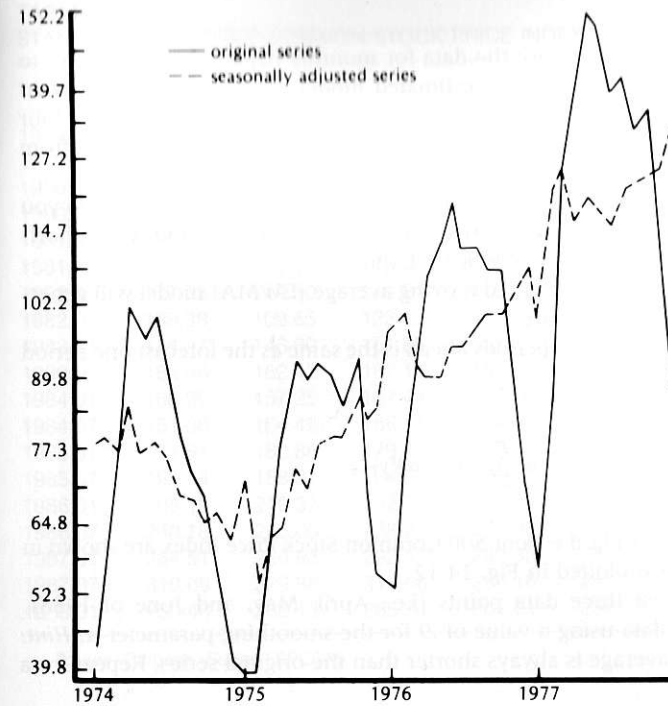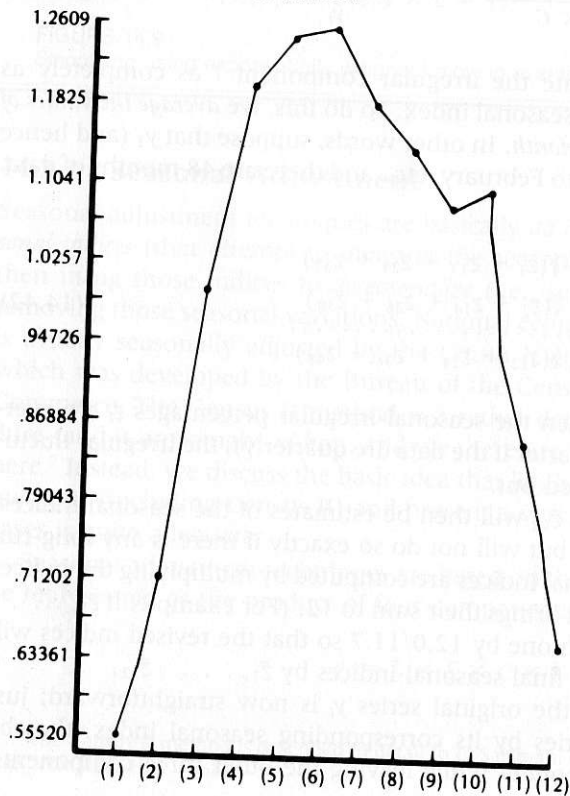
FIGURE 14.11
Seasonal adjustments of housing starts data.

Seasonal Indices

Month	Index	Month	Index
January	.5552	July	1.1900
February	.7229	August	1.1454
March	.9996	September	1.0675
April	1.1951	October	1.0823
May	1.2562	November	.8643
June	1.2609	December	.6584

These seasonal indices have been plotted in Fig. 14.10.

To deseasonalize the original series y_t we just divide each value in the series by its corresponding seasonal index, thereby removing the seasonal component. The original series y_t together with the seasonally adjusted series y_t^a are shown in Fig. 14.11. Observe that the seasonal variation has been eliminated in the adjusted series, while the long-run trend and short-run irregular fluctuations remain.

EXERCISES

14.1 Go back to Example 14.1 and use the data for monthly department store sales to estimate a quadratic trend model. Use the estimated model to obtain an extrapolated value for sales for April 1974. Try to evaluate your model in comparison to the other four estimated in Example 14.1, and explain how and why your forecast for April differs from the other forecasts in the example.

14.2 Which (if any) of the simple extrapolation models presented in Section 14.1 do you think might be suitable for forecasting the GNP? The Consumer Price Index? A short-term interest rate? Annual production of wheat? Explain.

14.3 Show that the exponentially weighted moving average (EWMA) model will generate forecasts that are adaptive in nature.

14.4 Show that the EWMA forecast l periods ahead is the same as the forecast one period ahead, i.e.,

$$\hat{y}_{T+l} = \alpha \sum_{\tau=0}^{\infty} (1 - \alpha)^{\tau} y_{T-\tau}$$

14.5 Monthly data for the Standard & Poor 500 Common Stock Price Index are shown in Table 14.1. The data are also plotted in Fig. 14.12.

(*a*) Using all but the last three data points (i.e., April, May, and June of 1988), exponentially smooth the data using a value of .9 for the smoothing parameter α. *Hint:* Remember that a moving average is always shorter than the original series. Repeat for a value of .2.

(*b*) Again using all but the last three data points, smooth the data using Holt's two-parameter exponential smoothing method. Set $\alpha = .2$ and $\gamma = .2$. Explain how and why the results differ from those in (*a*) above. Now use Eq. (14.36) to forecast the series out 1, 2, and 3 months. How close is your forecast to the actual values of the S&P 500 index for April to June 1988?

14.6 Monthly data for retail auto sales are shown in Table 14.2 on page 438. The data are also plotted in Fig. 14.13.

(*a*) Use a 6-month centered moving average to smooth the data. Is a seasonal pattern evident? Would you expect auto sales to exhibit seasonal regularities?

(*b*) Using the original data in Table 14.2, apply the seasonal adjustment procedure described in the text. Plot the 12 final seasonal indices as a function of time and try to explain the shape of the curve. Also plot the seasonally adjusted series and compare it to the original series.

TABLE 14.1
STANDARD & POOR 500 COMMON STOCK PRICE INDEX

Obs.						
1979.01	99.71	98.23	100.11	102.07	99.73	101.73
1979.07	102.71	107.36	108.60	104.47	103.66	107.78
1980.01	110.87	115.34	104.69	102.97	107.69	114.55
1980.07	119.83	123.50	126.51	130.22	135.65	133.48
1981.01	132.97	128.40	133.19	134.43	131.73	132.28
1981.07	129.13	129.63	118.27	119.80	122.92	123.79
1982.01	117.28	114.50	110.84	116.31	116.35	109.70
1982.07	109.38	109.65	122.43	132.66	138.10	139.37
1983.01	144.27	146.80	151.88	157.71	164.10	166.39
1983.07	166.96	162.42	167.16	167.65	165.23	164.36
1984.01	166.39	157.25	157.44	157.60	156.55	153.12
1984.07	151.08	164.42	166.11	164.82	166.27	164.48
1985.01	171.61	180.88	179.42	180.62	184.90	188.89
1985.07	192.54	188.31	184.06	186.18	197.45	207.26
1986.01	208.19	219.37	232.33	237.98	238.46	245.30
1986.07	240.18	245.00	238.27	237.36	245.09	248.61
1987.01	264.51	280.93	292.47	289.32	289.12	301.38
1987.07	310.09	329.36	318.66	280.16	245.01	240.96
1988.01	250.48	258.13	265.74	262.61	256.12	270.68

Source: Citibase, Series FSPCOM.

FIGURE 14.12
Standard & Poor 500 Common Stock Price Index.
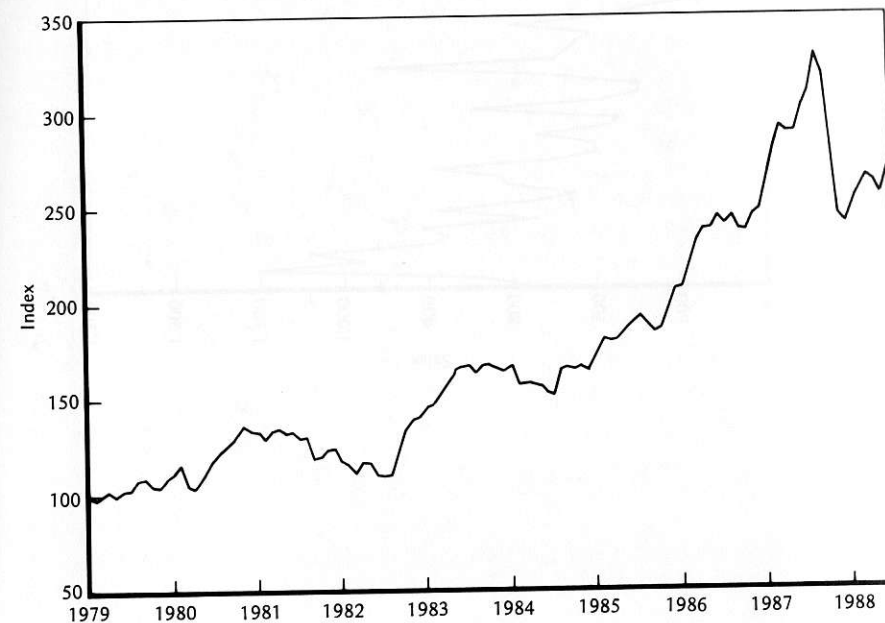
TABLE 14.2
RETAIL AUTO SALES (thousands of units)

Obs.						
1979.01	774.00	832.00	1,104.00	976.00	1,042.00	894.00
1979.07	876.00	908.00	767.00	892.00	768.00	726.00
1980.01	806.00	812.00	895.00	743.00	697.00	702.00
1980.07	773.00	686.00	672.00	848.00	698.00	649.00
1981.01	648.00	764.00	963.00	751.00	734.00	724.00
1981.07	707.00	801.00	687.00	649.00	585.00	523.00
1982.01	535.00	632.00	777.00	669.00	774.00	651.00
1982.07	630.00	609.00	671.00	656.00	743.00	632.00
1983.01	596.00	628.00	821.00	762.00	837.00	904.00
1983.07	792.00	741.00	705.00	861.00	782.00	752.00
1984.01	778.00	841.00	964.00	896.00	1,047.00	958.00
1984.07	890.00	814.00	744.00	900.00	802.00	759.00
1985.01	835.00	839.00	970.00	988.00	1,075.00	925.00
1985.07	899.00	1,001.00	1,068.00	864.00	762.00	812.00
1986.01	870.00	832.00	897.00	972.00	1,072.00	1,001.00
1986.07	954.00	952.00	1,217.00	906.00	783.00	992.00
1987.01	626.00	781.00	936.00	938.00	887.00	943.00
1987.07	913.00	968.00	905.00	802.00	737.00	843.00
1988.01	765.00	888.00	1,006.00	901.00	974.00	1,010.00

Source: Citibase, Series RCAR6T.
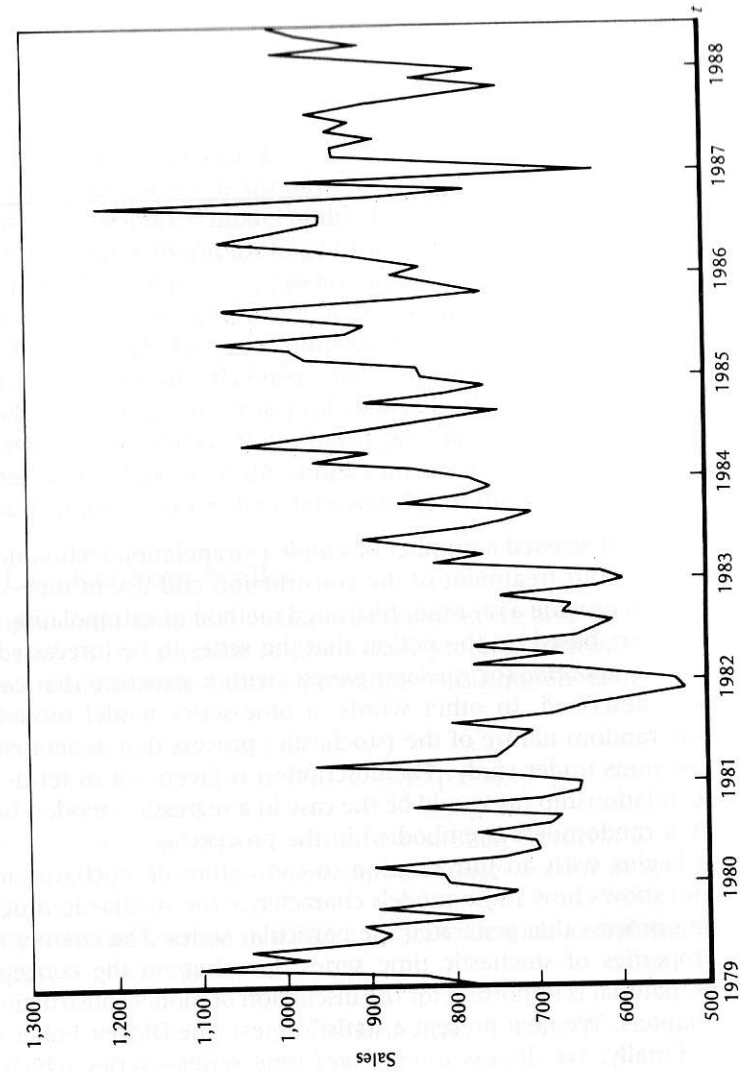
FIGURE 14.13
Monthly auto sales (thousands of units).

PROPERTIES OF STOCHASTIC TIME SERIES

In the last chapter we discussed a number of simple extrapolation techniques. In this chapter we begin our treatment of the construction and use of time-series models. Such models provide a more sophisticated method of extrapolating time series, in that they are based on the notion that the series to be forecasted has been generated by a *stochastic* (or *random*) *process*, with a structure that can be characterized and described. In other words, a time-series model provides a description of the random nature of the (stochastic) process that generated the sample of observations under study. The description is given not in terms of a cause-and-effect relationship (as would be the case in a regression model) but in terms of how that randomness is embodied in the process.

This chapter begins with an introduction to the nature of stochastic time-series models and shows how those models characterize the stochastic structure of the underlying process that generated the particular series. The chapter then turns to the properties of stochastic time series, focusing on the concept of *stationarity*. This material is important for the discussion of model construction in the following chapters. We next present a statistical test (the Dickey-Fuller test) for stationarity. Finally, we discuss *co-integrated* time series—series which are nonstationary but can be combined to form a stationary series.

15.1 INTRODUCTION TO STOCHASTIC TIME-SERIES MODELS

The time-series models developed in this and the following chapters are all based on an important assumption—that the series to be forecasted has been generated by a *stochastic process*. In other words, we assume that each value $y_1, y_2,$

. . . , y_T in the series is drawn randomly from a probability distribution. In modeling such a process, we attempt to describe the characteristics of its randomness. This should help us to infer something about the probabilities associated with alternative future values of the series.

To be completely general, we could assume that the observed series $y_1, . . . , y_T$ is drawn from a set of *jointly distributed random variables*. If we could somehow numerically specify the probability distribution function for our series, then we could actually determine the probability of one or another future outcome.

Unfortunately, the complete specification of the probability distribution function for a time series is usually impossible. However, it usually is possible to construct a simplified model of the time series which explains its randomness in a manner that is useful for forecasting purposes. For example, we might believe that the values of $y_1, . . . , y_T$ are normally distributed and are correlated with each other according to a simple first-order autoregressive process. The actual distribution might be more complicated, but this simple model may be a reasonable approximation. Of course, the usefulness of such a model depends on how closely it captures the true probability distribution and thus the true random behavior of the series. Note that it *need not* (and usually will not) match the actual past behavior of the series *since the series and the model are stochastic*. It should simply capture the characteristics of the series' randomness.

15.1.1 Random Walks

Our first (and simplest) example of a stochastic time series is the *random walk* process.[1] In the simplest random walk process, each successive *change* in y_t is drawn *independently* from a probability distribution with 0 mean. Thus, y_t is determined by

$$y_t = y_{t-1} + \varepsilon_t \qquad (15.1)$$

with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$. Such a process could be generated by successive flips of a coin, where a head receives a value of $+1$ and a tail receives a value of -1.

Suppose we wanted to make a forecast for such a random walk process. The forecast is given by

$$\hat{y}_{T+1} = E(y_{T+1}|y_T, . . . , y_1) \qquad (15.2)$$

But $y_{T+1} = y_T + \varepsilon_{T+1}$ is independent of $y_{T-1}, . . . , y_1$. Thus, the forecast one period ahead is simply

$$\hat{y}_{T+1} = y_T + E(\varepsilon_{T+1}) = y_T \qquad (15.3)$$

[1] The random walk process has often been used as a model for the movement of stock market prices. See, for example, E. F. Fama, "Random Walks in Stock Market Prices," *Financial Analysts Journal*, September-October 1965.

The forecast two periods ahead is

$$\hat{y}_{T+2} = E(y_{T+2}|y_T, \dots, y_1) = E(y_{T+1} + \varepsilon_{T+2})$$
$$= E(y_T + \varepsilon_{T+1} + \varepsilon_{T+2}) = y_T \tag{15.4}$$

Similarly, the forecast l periods ahead is also y_T.

Although the forecast $\hat{y}_{T+l}$ will be the same no matter how large l is, the variance of the forecast error will grow as l becomes larger. For the one-period forecast, the forecast error is given by

$$e_1 = y_{T+1} - \hat{y}_{T+1}$$
$$= y_T + \varepsilon_{T+1} - y_T = \varepsilon_{T+1} \tag{15.5}$$

and its variance is just $E(\varepsilon_{T+1}^2) = \sigma_\varepsilon^2$. For the two-period forecast,

$$e_2 = y_{T+2} - \hat{y}_{T+2}$$
$$= y_T + \varepsilon_{T+1} + \varepsilon_{T+2} - y_T = \varepsilon_{T+1} + \varepsilon_{T+2} \tag{15.6}$$

and its variance is

$$E[(\varepsilon_{T+1} + \varepsilon_{T+2})^2] = E(\varepsilon_{T+1}^2) + E(\varepsilon_{T+2}^2) + 2E(\varepsilon_{T+1}\varepsilon_{T+2}) \tag{15.7}$$

Since ε_{T+1} and ε_{T+2} are independent, the third term in Eq. (15.7) is 0 and the error variance is $2\sigma_\varepsilon^2$. Similarly, for the l-period forecast, the error variance is $l\sigma_\varepsilon^2$. Thus, the *standard error of forecast* increases with the square root of l. We can thus obtain *confidence intervals* for our forecasts, and these intervals will become wider as the forecast horizon increases. This is illustrated in Fig. 15.1. Note that the forecasts are all equal to the last observation y_T, but the confidence intervals represented by 1 standard deviation in the forecast error increase as the square root of l.

FIGURE 15.1
Forecasting a random walk.
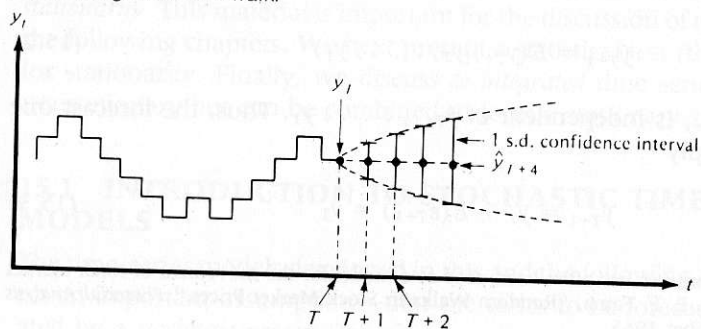
The fact that we can generate confidence intervals of this sort is an important advantage of stochastic time-series models. As we explained in Chapter 8, policy makers need to know the margin of error that has to be associated with a particular forecast, so confidence intervals can be as important as the forecasts themselves.

A simple extension of the random walk process discussed above is the random walk with drift. This process accounts for a trend (upward or downward) in the series y_t and thereby allows us to embody that trend in our forecast. In this process, y_t is determined by

$$y_t = y_{t-1} + d + \varepsilon_t \tag{15.8}$$

so that on the average the process will tend to move upward (for $d > 0$). Now the one-period forecast is

$$\hat{y}_{T+1} = E(y_{T+1}|y_T, \dots, y_1) = y_T + d \tag{15.9}$$

and the l-period forecast is

$$\hat{y}_{T+l} = y_T + ld \tag{15.10}$$

The standard error of forecast will be the same as before. For one period,

$$e_1 = y_{T+1} - \hat{y}_{T+1}$$
$$= y_T + d + \varepsilon_{T+1} - y_T - d = \varepsilon_{T+1} \tag{15.11}$$

as before. The process, together with forecasts and forecast confidence intervals, is illustrated in Fig. 15.2. As can be seen in that figure, the forecasts increase linearly with l, and the standard error of forecast increases with the square root of l.

In the next chapter we examine a general class of stochastic time-series models. Later, we will see how that class of models can be used to make forecasts for a wide variety of time series. First, however, it is necessary to introduce some basic concepts about stochastic processes and their properties.

15.1.2 Stationary and Nonstationary Time Series

As we begin to develop models for time series, we want to know whether or not the underlying stochastic process that generated the series can be assumed to be *invariant with respect to time*. If the characteristics of the stochastic process change over time, i.e., if the process is *nonstationary*, it will often be difficult to represent the time series over past and future intervals of time by a simple algebraic model.[2] On the other hand, if the stochastic process is fixed in time, i.e., if it is

[2] The random walk with drift is one example of a nonstationary process for which a simple forecasting model can be constructed.
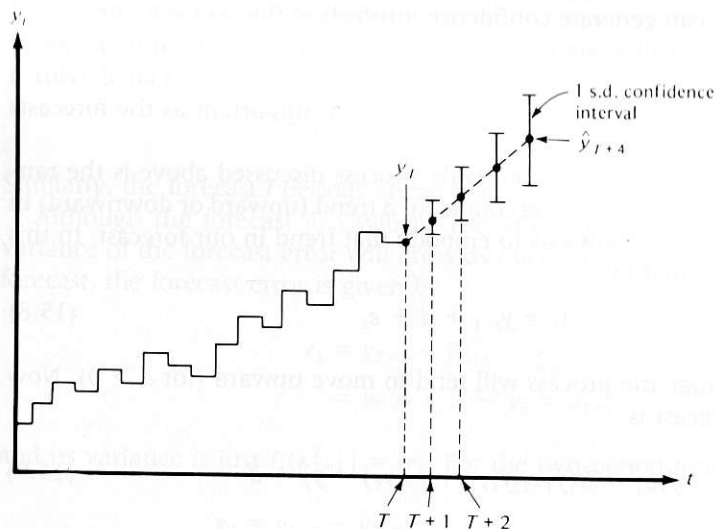
FIGURE 15.2
Forecasting a random walk with drift.

stationary, then one can model the process via an equation with fixed coefficients that can be estimated from past data. This is analogous to the single-equation regression model in which one economic variable is related to other economic variables, with coefficients that are estimated under the assumption that the structural relationship described by the equation is invariant over time (i.e., is stationary). If the structural relationship changed over time, we could not apply the techniques of Chapter 8 in using a regression model to forecast.

The models developed in detail in the next chapter of the book represent stochastic processes that are assumed to be in equilibrium about a constant mean level. The probability of a given fluctuation in the process from that mean level is assumed to be the same *at any point in time*. In other words, the stochastic properties of the stationary process are assumed to be *invariant with respect to time*.

One would suspect that many of the time series that one encounters in business and economics are not generated by stationary processes. The GNP, for example, has for the most part been growing steadily, and for this reason alone its stochastic properties in 1980 are different from those in 1933. Although it can be difficult to model nonstationary processes, we will see that nonstationary processes can often be transformed into stationary or approximately stationary processes.

15.1.3 Properties of Stationary Processes

We have said that any stochastic time series $y_1, \ldots, y_T$ can be thought of as having been generated by a set of jointly distributed random variables; i.e., the

set of data points $y_1, \ldots, y_T$ represents a particular outcome of the joint probability distribution function $p(y_1, \ldots, y_T)$.[3] Similarly, a *future* observation y_{T+1} can be thought of as being generated by a *conditional probability distribution function* $p(y_{T+1}|y_1, \ldots, y_T)$, that is, a probability distribution for y_{T+1} given the past observations $y_1, \ldots, y_T$. We define a *stationary* process, then, as one whose joint distribution and conditional distribution both are *invariant with respect to displacement in time*. In other words, if the series y_t is stationary, then

$$p(y_t, \ldots, y_{t+k}) = p(y_{t+m}, \ldots, y_{t+k+m}) \tag{15.12}$$

and

$$p(y_t) = p(y_{t+m}) \tag{15.13}$$

for any t, k, and m.

Note that if the series y_t is stationary, the *mean* of the series, defined as

$$\mu_y = E(y_t) \tag{15.14}$$

must also be stationary, so that $E(y_t) = E(y_{t+m})$, for any t and m. Furthermore, the *variance* of the series,

$$\sigma_y^2 = E[(y_t - \mu_y)^2] \tag{15.15}$$

must be stationary, so that $E[(y_t - \mu_y)^2] = E[y_{t+m} - \mu_y)^2]$, and finally, for any lag k, the *covariance* of the series,

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)] \tag{15.16}$$

must be stationary, so that $\text{Cov}(y_t, y_{t+k}) = \text{Cov}(y_{t+m}, y_{t+m+k})$.[4]

If a stochastic process is stationary, the probability distribution $p(y_t)$ is the same for all time t and its shape (or at least some of its properties) can be inferred by looking at a histogram of the observations $y_1, \ldots, y_T$ that make up the observed series. Also, an estimate of the mean μ_y of the process can be obtained from the *sample mean* of the series

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{15.17}$$

---

[3] This outcome is called a *realization*. Thus $y_1, \ldots, y_T$ represent one particular realization of the stochastic process represented by the probability distribution $p(y_1, \ldots, y_T)$.

[4] It is possible for the mean, variance, and covariances of the series to be stationary but not the joint probability distribution. If the probability distributions are stationary, we term the series *strict-sense stationary*. If the mean, variance, and covariances are stationary, we term the series *wide-sense stationary*. Note that strict-sense stationarity implies wide-sense stationarity but that the converse is not true.

and an estimate of the variance σ_y^2 can be obtained from the *sample variance*

$$\hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2 \qquad (15.18)$$

15.2 CHARACTERIZING TIME SERIES: THE AUTOCORRELATION FUNCTION

While it is usually impossible to obtain a complete description of a stochastic process (i.e., actually specify the underlying probability distributions), the autocorrelation function is extremely useful because it provides a partial description of the process for modeling purposes. The autocorrelation function tells us how much correlation there is (and by implication how much interdependency there is) between neighboring data points in the series y_t. We define the *autocorrelation with lag k* as

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sqrt{E[(y_t - \mu_y)^2]E[(y_{t+k} - \mu_y)^2]}} = \frac{\text{Cov}(y_t, y_{t+k})}{\sigma_{y_t}\sigma_{y_{t+k}}} \qquad (15.19)$$

For a stationary process the variance at time t in the denominator of Eq. (15.19) is the same as the variance at time $t + k$; thus the denominator is just the variance of the stochastic process, and

$$\rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sigma_y^2} \qquad (15.20)$$

Note that the numerator of Eq. (15.20) is the covariance between y_t and y_{t+k}, γ_k, so that

$$\rho_k = \frac{\gamma_k}{\gamma_0} \qquad (15.21)$$

and thus $\rho_0 = 1$ for *any* stochastic process.

Suppose that the stochastic process is simply

$$y_t = \varepsilon_t \qquad (15.22)$$

where ε_t is an independently distributed random variable with zero mean. Then it is easy to see from Eq. (15.20) that the autocorrelation function for this process is given by $\rho_0 = 1$, $\rho_k = 0$ for $k > 0$. The process of Eq. (15.22) is called *white noise*, and there is no model that can provide a forecast any better than $\hat{y}_{T+l} = 0$

for all l. Thus if the autocorrelation function is zero (or close to zero) for all $k > 0$, there is little or no value in using a model to forecast the series.

Of course the autocorrelation function in Eq. (15.20) is purely theoretical, in that it describes a stochastic process for which we have only a limited number of observations. In practice, then, we must calculate an *estimate* of the autocorrelation function, called the *sample autocorrelation function*:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2} \qquad (15.23)$$

It is easy to see from their definitions that both the theoretical and estimated autocorrelation functions are symmetrical, i.e., that the correlation for a positive displacement is the same as that for a negative displacement, so that

$$\rho_k = \rho_{-k} \qquad (15.24)$$

Then, when plotting an autocorrelation function (i.e., plotting ρ_k for different values of k), one need consider only positive values of k.

It is often useful to determine whether a particular value of the sample autocorrelation function $\hat{\rho}_k$ is close enough to zero to permit assuming that the *true* value of the autocorrelation function ρ_k is indeed equal to zero. It is also useful to test whether *all* the values of the autocorrelation function for $k > 0$ are equal to zero. (If they are, we know that we are dealing with white noise.) Fortunately, simple statistical tests exist that can be used to test the hypothesis that $\rho_k = 0$ for a particular k or to test the hypothesis that $\rho_k = 0$ for all $k > 0$.

To test whether a particular value of the autocorrelation function ρ_k is equal to zero we use a result obtained by Bartlett. He showed that if a time series has been generated by a *white noise* process, the sample autocorrelation coefficients (for $k > 0$) are approximately distributed according to a normal distribution with mean 0 and standard deviation $1/\sqrt{T}$ (where T is the number of observations in the series).[5] Thus, if a particular series consists of, say, 100 data points, we can attach a standard error of .1 to each autocorrelation coefficient. Therefore, if a particular coefficient was greater in magnitude than .2, we could be 95 percent sure that the true autocorrelation coefficient is not zero.

To test the *joint hypothesis* that *all* the autocorrelation coefficients are zero we use the Q statistic introduced by Box and Pierce. We will discuss this statistic in some detail in Chapter 17 in the context of performing diagnostic checks on

[5] See M. S. Bartlett, "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society*, ser. B8, vol. 27, 1946. Also see G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (San Francisco: Holden-Day, 1970).

estimated time-series models, so here we only mention it in passing. Box and Pierce show that the statistic

$$Q = T \sum_{k=1}^{K} \hat{\rho}_k^2 \qquad (15.25)$$

is (approximately) distributed as chi square with K degrees of freedom. Thus if the calculated value of Q is greater than, say, the critical 5 percent level, we can be 95 percent sure that the *true* autocorrelation coefficients $\rho_1, \ldots, \rho_k$ are not all zero.

In practice people tend to use the critical 10 percent level as a cutoff for this test. For example, if Q turned out to be 18.5 for a total of $K = 15$ lags, we would observe that this is below the critical level of 22.31 and accept the hypothesis that the time series was generated by a white noise process.

Let us now turn to an example of an estimated autocorrelation function for a stationary economic time series. We have calculated $\hat{\rho}_k$ for quarterly data on real nonfarm inventory investment (measured in billions of 1982 dollars). The time series itself (covering the period 1952 through the first two quarters of 1988) is shown in Fig. 15.3, and the sample autocorrelation function is shown in Fig. 15.4. Note that the autocorrelation function falls off rather quickly as the lag k increases. This is typical of a stationary time series, such as inventory investment.

FIGURE 15.3
Nonfarm inventory investment (in 1982 constant dollars).
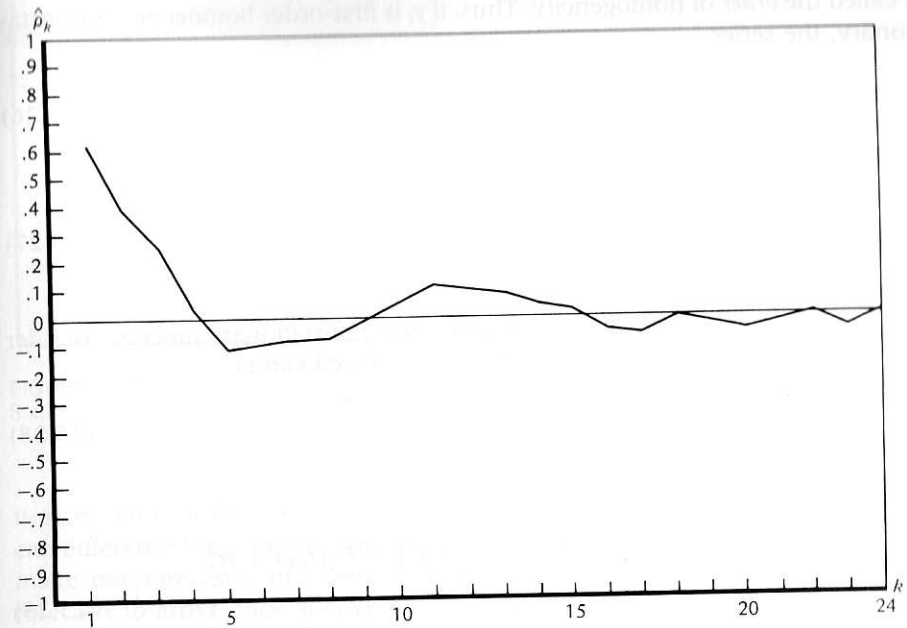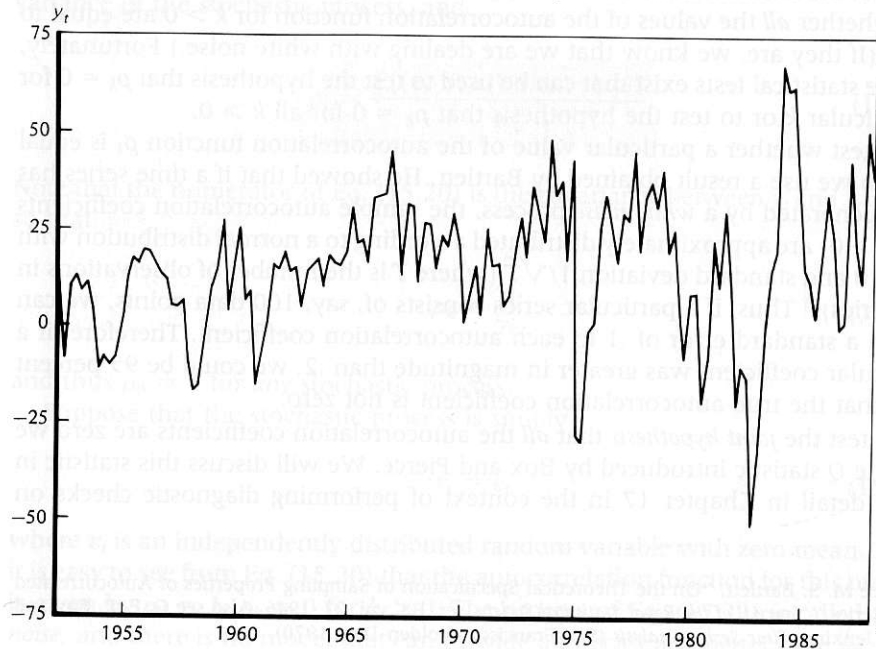
FIGURE 15.4
Nonfarm inventory investment: sample autocorrelation function.

In fact, as we will see, the autocorrelation function can be used to test whether or not a series is stationary. If $\hat{\rho}_k$ does not fall off quickly as k increases, this is an indication of nonstationarity. We will discuss more formal tests of nonstationarity ("unit root" tests) in Section 15.3.

If a time series is stationary, there exist certain analytical conditions which place bounds on the values that can be taken by the individual points of the autocorrelation function. However, the derivation of these conditions is somewhat complicated and will not be presented at this point. Furthermore, the conditions themselves are rather cumbersome and of limited usefulness in applied time-series modeling. Therefore, we have relegated them to Appendix 15.1. We turn our attention now to the properties of those time series which are nonstationary but which can be transformed into stationary series.

15.2.1 Homogeneous Nonstationary Processes

Probably very few of the time series one meets in practice are stationary. Fortunately, however, many of the nonstationary time series encountered (and this includes most of those that arise in economics and business) have the desirable property that if they are *differenced one or more times, the resulting series will be stationary*. Such a nonstationary series is termed *homogeneous*. The number of times that the original series must be differenced before a stationary series results

is called the *order* of homogeneity. Thus, if y_t is first-order homogeneous nonstationary, the series

$$w_t = y_t - y_{t-1} = \Delta y_t \qquad (15.26)$$

is stationary. If y_t happened to be second-order homogeneous, the series

$$w_t = \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \qquad (15.27)$$

would be stationary.

As an example of a first-order homogeneous nonstationary process, consider the simple random walk process that we introduced earlier:

$$y_t = y_{t-1} + \varepsilon_t \qquad (15.28)$$

Let us examine the variance of this process:

$$\gamma_0 = E(y_t^2) = E[(y_{t-1} + \varepsilon_t)^2] = E(y_{t-1}^2) + \sigma_\varepsilon^2$$
$$= E(y_{t-2}^2) + 2\sigma_\varepsilon^2 \qquad (15.29)$$
$$\vdots$$

or

$$\gamma_0 = E(y_{t-n}^2) + n\sigma_\varepsilon^2 \qquad (15.30)$$

Observe from this recursive relation that the variance is infinite and hence undefined. The same is true for the covariances, since, for example,

$$\gamma_1 = E(y_t y_{t-1}) = E[y_{t-1}(y_{t-1} + \varepsilon_t)] = E(y_{t-1}^2) \qquad (15.31)$$

Now let us look at the series that results from differencing the random walk process, i.e., the series

$$w_t = \Delta y_t = y_t - y_{t-1} = \varepsilon_t \qquad (15.32)$$

Since the ε_t are assumed independent over time, w_t is clearly a stationary process. Thus, we see that the random walk process is first-order homogeneous. In fact, w_t is just a white noise process, and it has the autocorrelation function $\rho_0 = 1$, but $\rho_k = 0$ for $k > 0$.

15.2.2 Stationarity and the Autocorrelation Function

The GNP or a series of sales figures for a firm are both likely to be nonstationary. Each has been growing (on average) over time, so that the mean of each series is time-dependent. It is quite likely, however, that if the GNP or company sales figures are first-differenced one or more times, the resulting series will be sta-
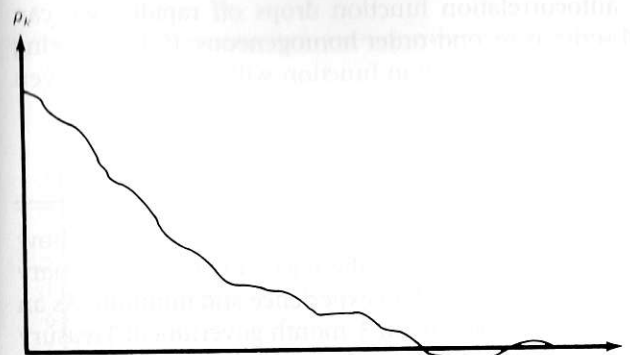
FIGURE 15.5
Stationary series.

tionary. Thus, if we want to build a time-series model to forecast the GNP, we can difference the series one or two times, construct a model for this new series, make our forecasts, and then *integrate* (i.e., undifference) the model and its forecasts to arrive back at GNP.

How can we decide whether a series is stationary or determine the appropriate number of times a homogeneous nonstationary series should be differenced to arrive at a stationary series? We can begin by looking at a plot of the autocorrelation function (called a *correlogram*). Figures 15.5 and 15.6 show autocorrelation functions for stationary and nonstationary series. The autocorrelation function for a stationary series drops off as k, the number of lags, becomes large, but this is usually not the case for a nonstationary series. If we are differencing a nonstationary series, we can test each succeeding difference by looking at the autocorrelation function. If, for example, the second round of differencing

FIGURE 15.6
Nonstationary series.
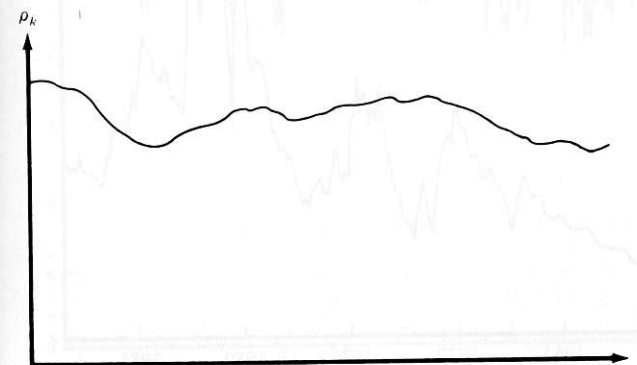
results in a series whose autocorrelation function drops off rapidly, we can determine that the original series is second-order homogeneous. If the resulting series is still nonstationary, the autocorrelation function will remain large even for long lags.

Example 15.1 Interest Rate   Often in applied work it is not clear how many times a nonstationary series should be differenced to yield a stationary one, and one must make a judgment based on experience and intuition. As an example, we will examine the interest rate on 3-month government Treasury bills. This series, consisting of monthly data from the beginning of 1950 through June 1988, is shown in Fig. 15.7, and its autocorrelation function is shown in Fig. 15.8. The autocorrelation function does decline as the number of lags becomes large, but only very slowly. In addition, the series exhibits an upward trend (so that the mean is not constant over time). We would therefore suspect that this series has been generated by a homogeneous nonstationary process. To check, we difference the series and recalculate the sample autocorrelation function.

The differenced series is shown in Fig. 15.9. Note that the mean of the series is now about constant, although the variance becomes unusually high

FIGURE 15.8
Three-month Treasury bill rate: sample autocorrelation function.

FIGURE 15.9
Three-month Treasury bill rate—first differences.

FIGURE 15.7
Three-month Treasury bill rate.
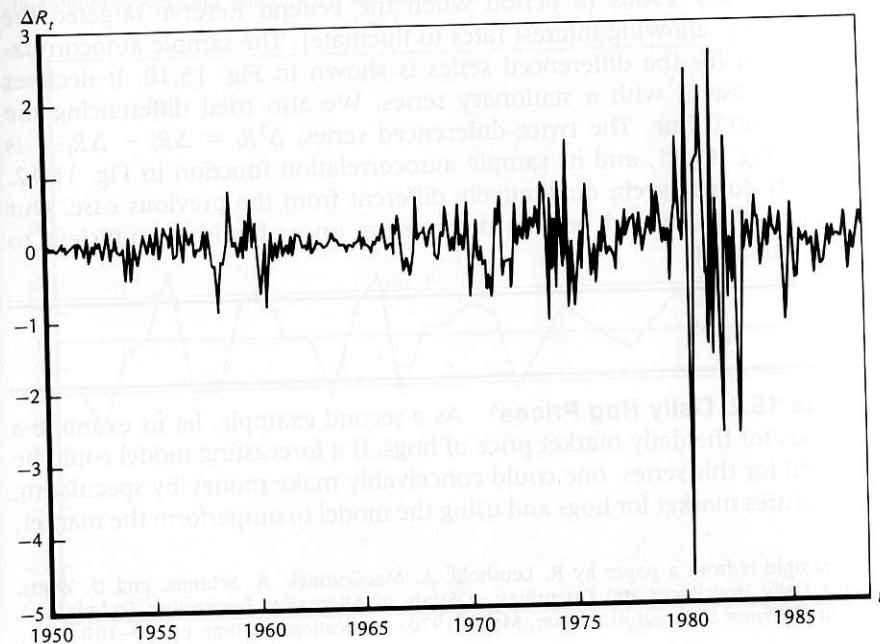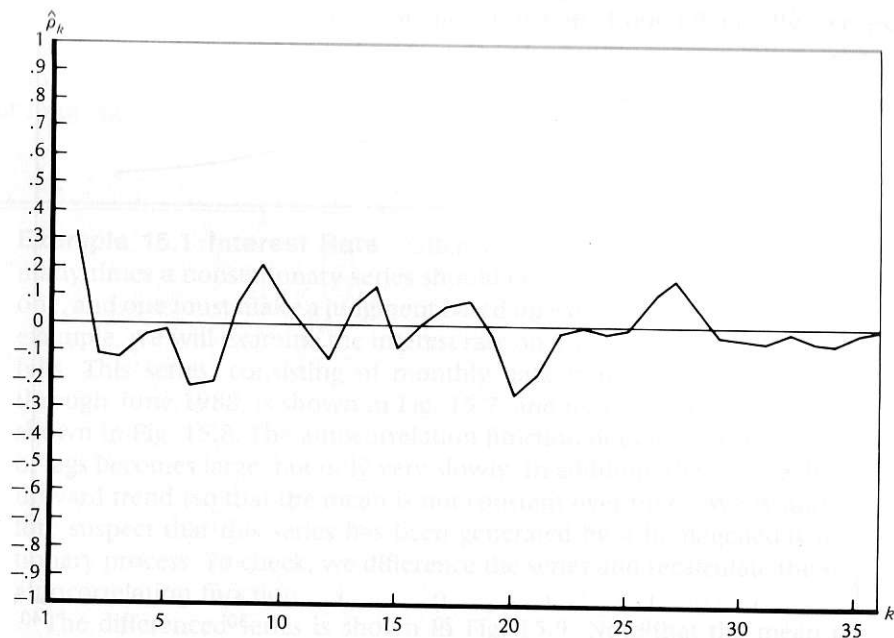
FIGURE 15.10
Interest rate—first differences: sample autocorrelation function.

FIGURE 15.11
Three-month Treasury bill rate—second differences.

FIGURE 15.12
Interest rate—second differences: sample autocorrelation function.

during the early 1980s (a period when the Federal Reserve targeted the money supply, allowing interest rates to fluctuate). The sample autocorrelation function for the differenced series is shown in Fig. 15.10. It declines rapidly, consistent with a stationary series. We also tried differencing the series a second time. The twice-differenced series, $\Delta^2 R_t = \Delta R_t - \Delta R_{t-1}$, is shown in Fig. 15.11, and its sample autocorrelation function in Fig. 15.12. The results do not seem qualitatively different from the previous case. Our conclusion, then, would be that differencing once should be sufficient to ensure stationarity.

Example 15.2 Daily Hog Prices[6]   As a second example, let us examine a time series for the daily market price of hogs. If a forecasting model could be developed for this series, one could conceivably make money by speculating on the futures market for hogs and using the model to outperform the market.

[6] This example is from a paper by R. Leuthold, A. MacCormick, A. Schmitz, and D. Watts, "Forecasting Daily Hog Prices and Quantities: A Study of Alternative Forecasting Techniques," *Journal of the American Statistical Association*, March 1970, Applications Section, pp. 90–107.

FIGURE 15.13
Sample autocorrelation functions of daily hog price data.

The series consists of 250 daily data points covering all the trading days in 1965. The price variable is the average price in dollars per hundredweight of all hogs sold in the eight regional markets in the United States on a particular day. The sample autocorrelation functions for the original price series and for the first difference of the series are shown in Fig. 15.13.

Observe that the original series is clearly nonstationary. The autocorrelation function barely declines, even after a 16-period lag. The series is, however, first-order homogeneous, since its first difference is clearly stationary.

In fact, not only is the first-differenced series stationary, but it appears to resemble white noise, since the sample autocorrelation function $\hat{\rho}_k$ is close to zero for all $k > 0$. To determine whether the differenced series is indeed white noise, let us calculate the Q statistic for the first 15 lags. The value of this statistic is 14.62, which, with 15 degrees of freedom, is insignificant at the 10 percent level. We can therefore conclude that the differenced series is white noise and that the original price series can best be modeled as a *random walk*:

$$P_t = P_{t-1} + \varepsilon_t \tag{15.33}$$

As is the case of most stock market prices, our best forecast of P_t is its most recent value, and (sadly) there is no model that can help us outperform the market.

15.2.3 Seasonality and the Autocorrelation Function

We have just seen that the autocorrelation function can reveal information about the stationarity of a time series. In the remaining chapters of this book we will see that other information about a time series can be obtained from its autocorrelation function. However, we continue here by examining the relationship between the autocorrelation function and the *seasonality* of a time series.

As discussed in the previous chapter, seasonality is just a cyclical behavior that occurs on a regular calendar basis. An example of a highly seasonal time series would be toy sales, which exhibit a strong peak every Christmas. Sales of ice cream and iced-tea mix show seasonal peaks each summer in response to increased demand brought about by warmer weather; Peruvian anchovy production shows seasonal troughs once every 7 years in response to decreased supply brought about by cyclical changes in the ocean currents.

Often seasonal peaks and troughs are easy to spot by direct observation of the time series. However, if the time series fluctuates considerably, seasonal peaks and troughs might not be distinguishable from the other fluctuations. Recognition of seasonality is important because it provides information about "regularity" in the series that can aid us in making a forecast. Fortunately, that recognition can be made easier with the help of the autocorrelation function.
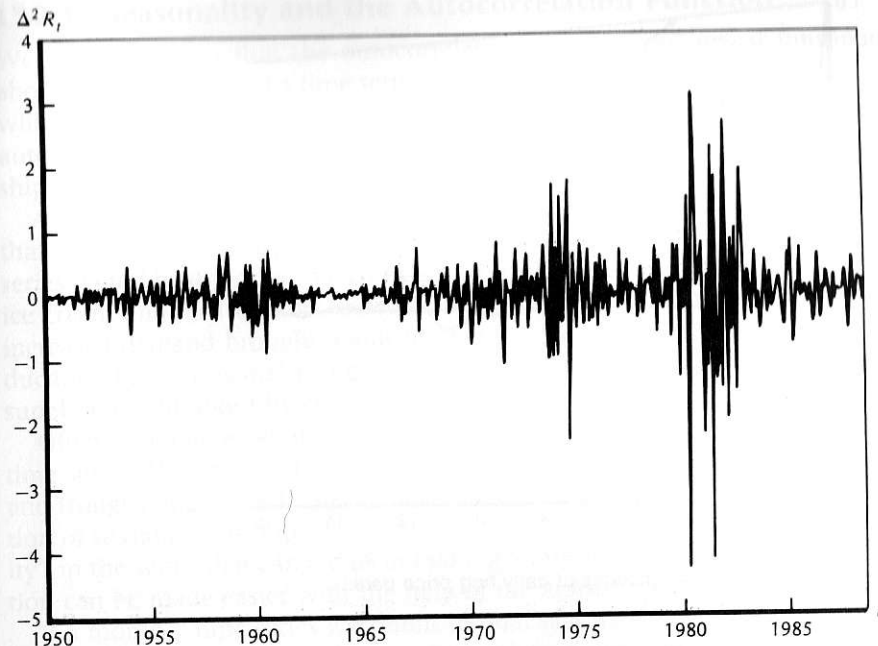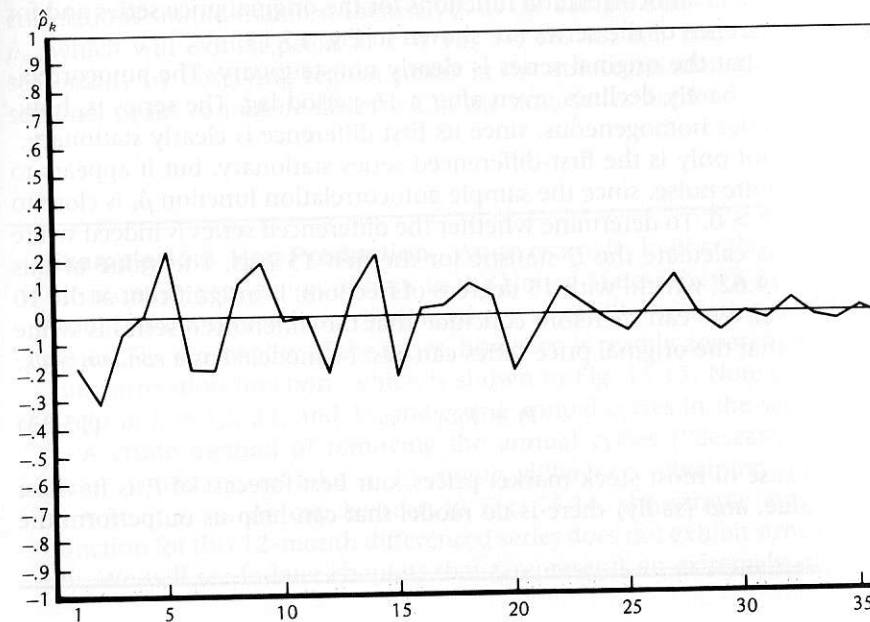
If a monthly time series y_t exhibits annual seasonality, the data points in the series should show some degree of correlation with the corresponding data points which lead or lag by 12 months. In other words, we would expect to see some degree of correlation between y_t and y_{t-12}. Since y_t and y_{t-12} will be correlated, as will y_{t-12} and y_{t-24}, we should also see correlation between y_t and y_{t-24}. Similarly there will be correlation between y_t and y_{t-36}, y_t and y_{t-48}, etc. These correlations should manifest themselves in the sample autocorrelation function $\hat{\rho}_k$, which will exhibit peaks at $k = 12, 24, 36, 48$, etc. Thus we can identify seasonality by observing regular peaks in the autocorrelation function, even if seasonal peaks cannot be discerned in the time series itself.

Example 15.3 Hog Production   As an example, look at the time series for the monthly production of hogs in the United States, shown in Fig. 15.14. It would take a somewhat experienced eye to easily discern seasonality in that series. The seasonality of the series, however, is readily apparent in its sample autocorrelation function, which is shown in Fig. 15.15. Note the peaks that occur at $k = 12, 24$, and 36, indicating annual cycles in the series.

A crude method of removing the annual cycles ("deseasonalizing" the data) would be to take a 12-month difference, obtaining a new series $z_t = y_t - y_{t-12}$. As can be seen in Fig. 15.16, the sample autocorrelation function for this 12-month differenced series does not exhibit strong seasonality. We will see in later chapters that z_t represents an extremely simple time-
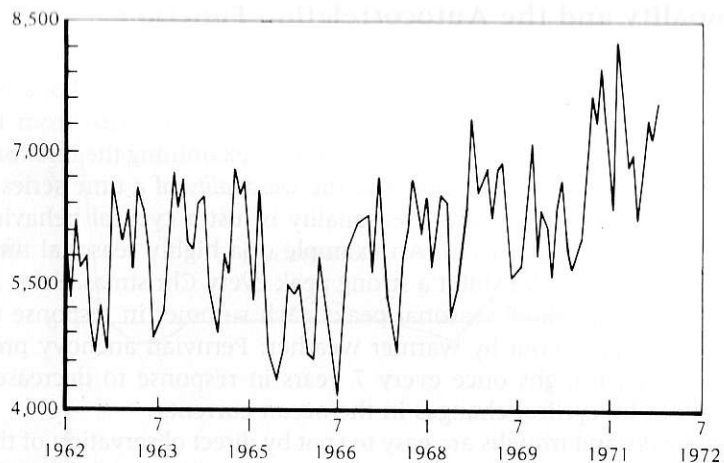
FIGURE 15.14
Hog production (in thousands of hogs per month). Time bounds:
January 1962 to December 1971.

series model for hog production, since it accounts only for the annual cycle.
We can complete this example by observing that the autocorrelation function
in Fig. 15.16 declines only slowly, so that there is some doubt as to whether z_t
is a stationary series. We therefore first-differenced this series, to obtain $w_t = \Delta z_t = \Delta(y_t - y_{t-12})$. The sample autocorrelation function of this series, shown
in Fig. 15.17, declines rapidly and remains small, so that we can be confident
that w_t is a stationary, nonseasonal time series.

FIGURE 15.15
Sample autocorrelation function for hog production series.
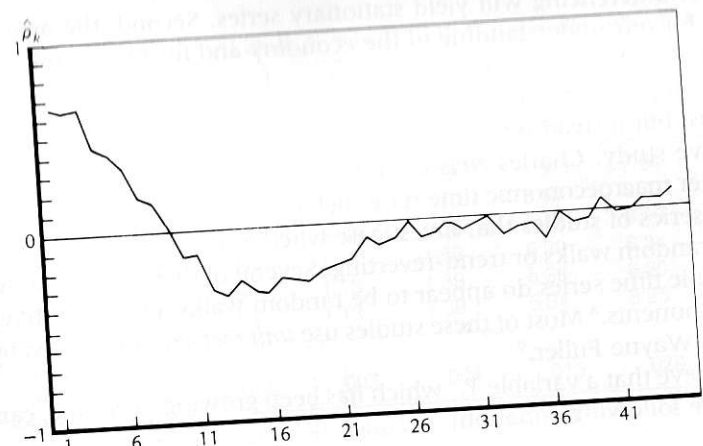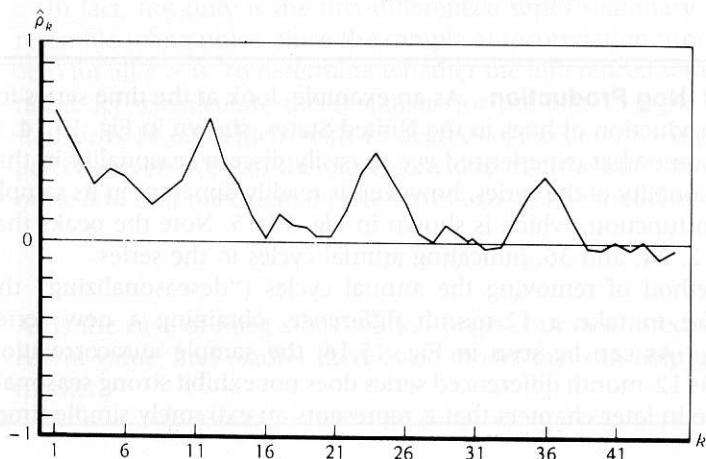
FIGURE 15.16
Hog production: sample autocorrelation function of $y_t - y_{t-12}$.
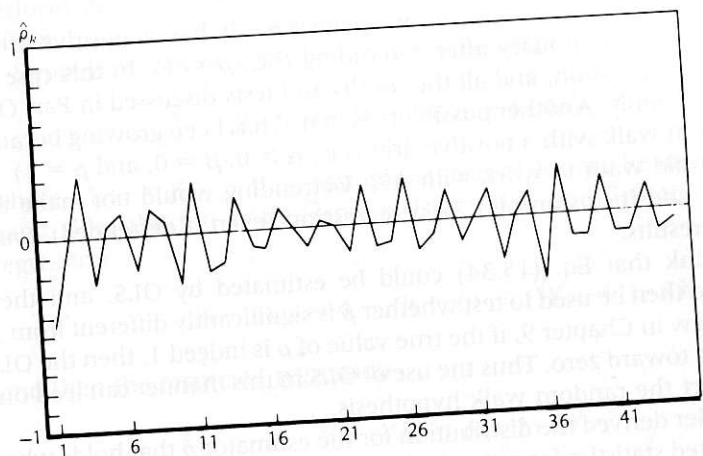
FIGURE 15.17
Hog production: sample autocorrelation function of $\Delta(y_t - y_{t-12})$.

15.3 TESTING FOR RANDOM WALKS

Do economic variables such as GNP, employment, and interest rates tend to
revert back to some long-run trend following a shock, or do they follow random
walks? This question is important for two reasons. First, if these variables follow
random walks, a regression of one against another can lead to spurious results.
[The Gauss-Markov theorem would not hold, for example, because a random
walk does not have a finite variance. Hence ordinary least squares (OLS) would
not yield a consistent parameter estimator.] Detrending the variables before
running the regression will not help; the detrended series will still be nonsta-

tionary. Only first-differencing will yield stationary series. Second, the answer has implications for our understanding of the economy and for forecasting. If a variable like GNP follows a random walk, the effects of a temporary shock (such as an increase in oil prices or a drop in government spending) will not dissipate after several years, but instead will be permanent.

In a provocative study, Charles Nelson and Charles Plosser found evidence that GNP and other macroeconomic time series behave like random walks.[7] The work spawned a series of studies that investigate whether economic and financial variables are random walks or trend-reverting. Several of these studies show that many economic time series do appear to be random walks, or at least have random walk components.[8] Most of these studies use *unit root tests* introduced by David Dickey and Wayne Fuller.[9]

Suppose we believe that a variable Y_t, which has been growing over time, can be described by the following equation:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t \qquad (15.34)$$

One possibility is that Y_t has been growing because it has a positive trend ($\beta > 0$), but would be stationary after detrending (i.e., $\rho < 1$). In this case, Y_t could be used in a regression, and all the results and tests discussed in Part One of this book would apply. Another possibility is that Y_t has been growing because it follows a random walk with a positive drift (i.e., $\alpha > 0$, $\beta = 0$, and $\rho = 1$). In this case, one would want to work with ΔY_t. Detrending would not make the series stationary, and inclusion of Y_t in a regression (even if detrended) could lead to spurious results.

One might think that Eq. (15.34) could be estimated by OLS, and the t statistic on $\hat{\rho}$ could then be used to test whether $\hat{\rho}$ is significantly different from 1. However, as we saw in Chapter 9, if the true value of ρ is indeed 1, then the OLS estimator is biased toward zero. Thus the use of OLS in this manner can lead one to incorrectly reject the random walk hypothesis.

Dickey and Fuller derived the distribution for the estimator $\hat{\rho}$ that holds when $\rho = 1$, and generated statistics for a simple F test of the random walk hypothesis, i.e., of the hypothesis that $\beta = 0$ and $\rho = 1$. The Dickey-Fuller test is easy to

[7] C. R. Nelson and C. I. Plosser, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, vol. 10, pp. 139–162, 1982.

[8] Examples of these studies include J. Y. Campbell and N. G. Mankiw, "Are Output Fluctuations Transitory?," *Quarterly Journal of Economics*, vol. 102, pp. 857–880, 1987; J. Y. Campbell and N. G. Mankiw, "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review Papers and Proceedings*, vol. 77, pp. 111–117, 1987; and G. W. Gardner and K. P. Kimbrough, "The Behavior of U.S. Tariff Rates," *American Economic Review*, vol. 79, pp. 211–218, 1989.

[9] D. A. Dickey and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time-Series with a Unit Root," *Journal of the American Statistical Association*, vol. 74, pp. 427–431, 1979; D. A. Dickey and W. A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, vol. 49, pp. 1057–1072, 1981; and W. A. Fuller, *Introduction to Statistical Time Series* (New York: Wiley, 1976).

TABLE 15.1
DISTRIBUTION OF F FOR $(\alpha, \beta, \rho) = (\alpha, 0, 1)$ IN $Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t$

Sample size N	Probability of a smaller value							
	.01	.025	.05	.10	.90	.95	.975	.99
25	.74	.90	1.08	1.33	5.91	7.24	8.65	10.61
50	.76	.93	1.11	1.37	5.61	6.73	7.81	9.31
100	.76	.94	1.12	1.38	5.47	6.49	7.44	8.73
250	.76	.94	1.13	1.39	5.39	6.34	7.25	8.43
500	.76	.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	.77	.94	1.13	1.39	5.34	6.25	7.16	8.27
Standard error	.004	.004	.003	.004	.015	.020	.032	.058

Source: Dickey and Fuller, op. cit., Table VI, p. 1063, 1981.

perform, and can be applied to a more general version of Eq. (15.34). It works as follows.

Suppose Y_t can be described by the following equation:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \lambda_1 \Delta Y_{t-1} + \varepsilon_t \qquad (15.35)$$

where $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$. (Additional lags of ΔY_t can be included on the right-hand side; the test is the same.) Using OLS, one first runs the unrestricted regression

$$Y_t - Y_{t-1} = \alpha + \beta t + (\rho - 1)Y_{t-1} + \lambda_1 \Delta Y_{t-1} \qquad (15.36)$$

and then the restricted regression

$$Y_t - Y_{t-1} = \alpha + \lambda_1 \Delta Y_{t-1} \qquad (15.37)$$

Then, one calculates the standard F ratio to test whether the restrictions ($\beta = 0$, $\rho = 1$) hold.[10] This ratio, however, is *not* distributed as a standard F distribution under the null hypothesis. Instead, one must use the distributions tabulated by Dickey and Fuller. Critical values for this statistic are shown in Table 15.1.

Note that these critical values are much larger than those in the standard F table. For example, if the calculated F ratio turns out to be 5.2 and there are 100

[10] Recall that F is calculated as follows:

$$F = (N - k)(\text{ESS}_R - \text{ESS}_{UR})/q(\text{ESS}_{UR})$$

where ESS_R and ESS_{UR} are the sums of squared residuals in the restricted and unrestricted regressions, respectively, N is the number of observations, k is the number of estimated parameters in the unrestricted regression, and q is the number of parameter restrictions.

observations, we would easily reject the null hypothesis of a unit root at the 5 percent level if we used a standard F table (which, with two parameter restrictions, shows a critical value of about 3.1), i.e., we would conclude that there is no random walk. This rejection, however, would be incorrect. Note that we fail to reject the hypothesis of a random walk using the distribution calculated by Dickey and Fuller (the critical value is 6.49).[11]

Although the Dickey-Fuller test is widely used, one should keep in mind that its power is limited. It only allows us to reject (or fail to reject) the hypothesis that a variable is *not* a random walk. A failure to reject (especially at a high significance level) is only weak evidence in favor of the random walk hypothesis.

Example 15.4 Do Commodity Prices Follow Random Walks? Like stocks and bonds, many commodities are actively traded in highly liquid spot markets. In addition, trading is active in financial instruments such as futures contracts that depend on the prices of these commodities. One might therefore expect the prices of these commodities to follow random walks, so that no investor could expect to profit by following some trading rule. (See Example 15.2 on daily hog prices.) Indeed, most financial models of futures, options, and other instruments tied to a commodity are based on the assumption that the spot price follows a random walk.[12]

On the other hand, basic microeconomic theory tells us that in the long run the price of a commodity ought to be tied to its marginal production cost. This means that although the price of a commodity might be subject to sharp short-run fluctuations, it ought to tend to return to a "normal" level based on cost. Of course, marginal production cost might be expected to slowly rise (if the commodity is a depletable resource) or fall (because of technological change), but that means the *detrended* price should tend to revert back to a normal level.

At issue, then, is whether the price of a commodity can best be described as a random walk process, perhaps with trend:

$$P_t = \alpha + P_{t-1} + \varepsilon_t \qquad (15.38)$$

where ε_t is a white noise error term, or alternatively as a first-order autoregressive process with trend:

$$P_t = \alpha + \beta t + \rho P_{t-1} + \varepsilon_t \qquad (15.39)$$

[11] For further discussion of the random walk model and alternative tests, see P. Perron, "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach," *Journal of Economic Dynamics and Control*, vol. 12, pp. 297–332, 1988, and P. C. B. Phillips, "Time Series Regression with Unit Roots," *Econometrica*, vol. 55, pp. 277–302, 1987.

[12] For a thorough treatment of commodity markets and derivative instruments such as futures contracts, see Darrell Duffie, *Futures Markets* (Englewood Cliffs, N.J.: Prentice-Hall, 1989), and John Hull, *Options, Futures, and Other Derivative Securities* (Englewood Cliffs, N.J.: Prentice-Hall, 1989).
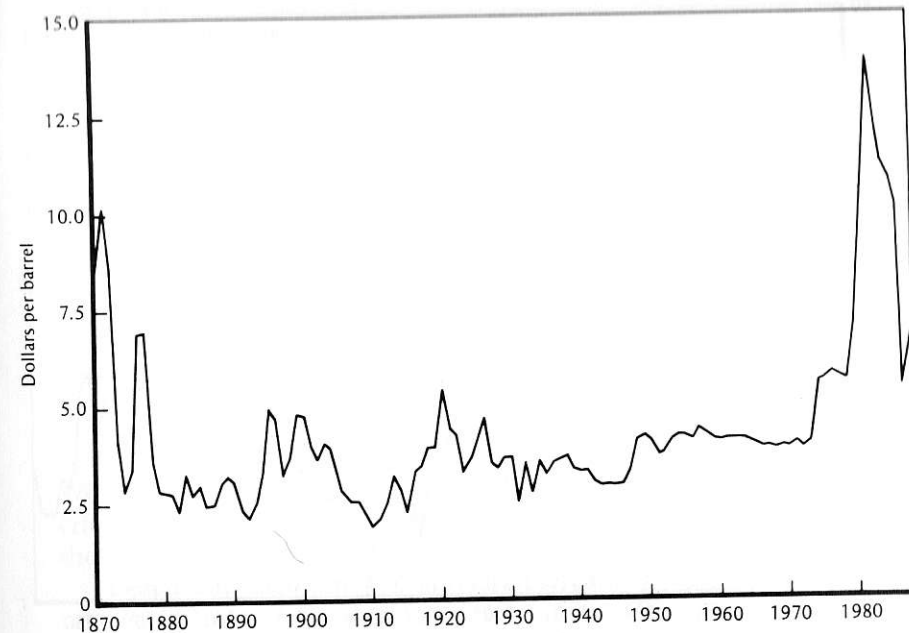
FIGURE 15.18
Price of oil (in 1967 constant dollars).

Since any reversion to long-run marginal cost is likely to be slow, we will only be able to discriminate between these two alternative models with data that cover a long time period (so that short-run fluctuations wash out). Fortunately, more than 100 years of commodity price data are available.

Figures 15.18, 15.19, and 15.20 show the real (in 1967 dollars) prices of crude oil, copper, and lumber over the 117-year period 1870 to 1987.[13] Observe that the price of oil fluctuated around $4 per barrel from 1880 to 1970 but rose sharply in 1974 and 1980–81 and then fell during the mid-1980s. Copper prices have fluctuated considerably but show a general downward trend, while lumber prices have tended to increase, at least up to about 1950.

We ran a Dickey-Fuller unit root test on each price series by estimating the unrestricted regression:

$$P_t - P_{t-1} = \alpha + \beta t + (\rho - 1)P_{t-1} + \lambda \, \Delta P_{t-1} + \varepsilon_t$$

and the restricted regression:

$$P_t - P_{t-1} = \alpha + \lambda \, \Delta P_{t-1} + \varepsilon_t$$

[13] The data for 1870 to 1973 are from Robert Manthy, *A Century of Natural Resource Statistics*, Johns Hopkins University Press, 1978. Data after 1973 are from publications of the Energy Information Agency and U.S. Bureau of Mines. All prices are deflated by the wholesale price index (now the Producer Price Index).
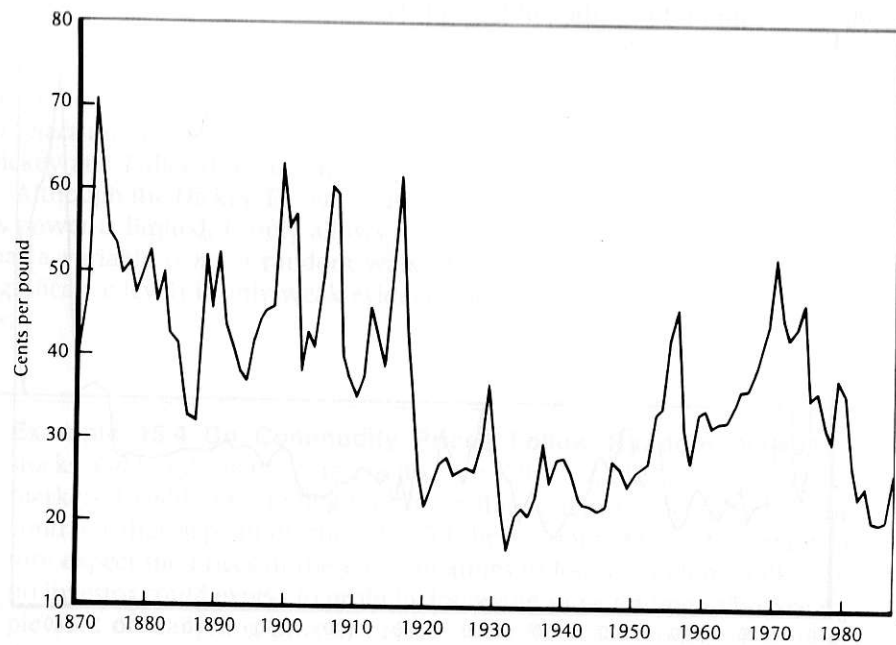
FIGURE 15.19
Price of copper (in 1967 constant dollars).

FIGURE 15.20
Price of lumber (in 1967 constant dollars).
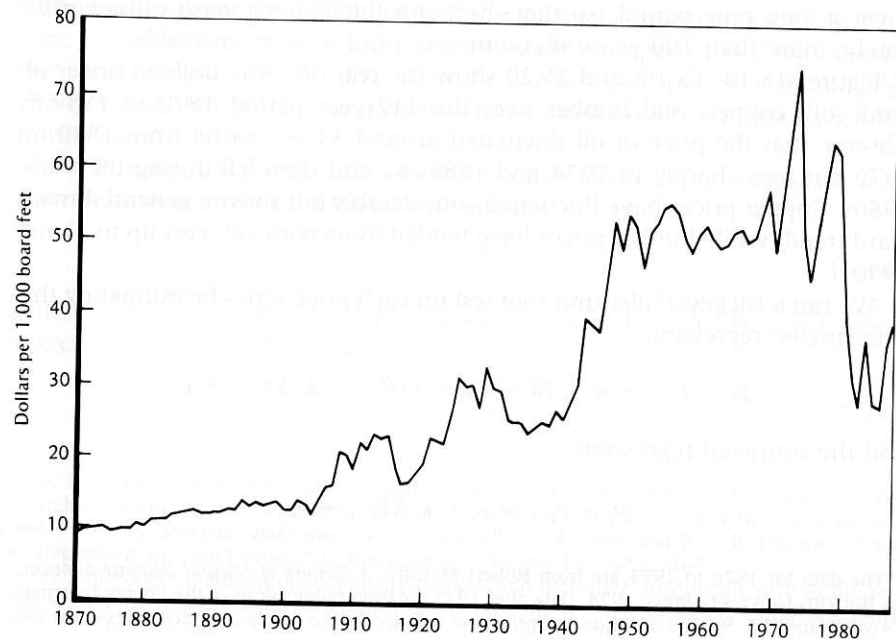
TABLE 15.2
DICKEY-FULLER TESTS

Commodity	α	β	$(\rho - 1)$	λ	ESS
Copper (unrestricted)	11.357	−.0446	−.2417	.0766	4,344.8
	(3.228)	(.0208)	(.0631)	(.0941)	
Copper (restricted)	−.1969			−.0440	4,913.5
	(.6098)			(.0934)	
Lumber (unrestricted)	.8825	.0660	−.1560	.1392	2,242.8
	(.8689)	(.0263)	(.0515)	(.0958)	
Lumber (restricted)	.2488			.0507	2,428.6
	(.4291)			(.0938)	
Oil (unrestricted)	.4366	.00895	−.2355	.2546	100.09
	(.2188)	(.00285)	(.0459)	(.0848)	
Oil (restricted)	−.0262			.1760	125.00
	(.0973)			(.0918)	

We tested the restrictions by calculating an F ratio and comparing it to the critical values in Table 15.1. Regression results (with standard errors in parentheses) are shown in Table 15.2.

In each case there are 116 annual observations. Hence, for copper the F ratio is $(112)(4,913.5 - 4,344.8)/(2)(4,344.8) = 7.33$. Comparing this to the critical values for a sample size of 100 in Table 15.1, we see that we can reject the hypothesis of a random walk at the 5 percent level. For lumber, the F ratio is 4.64, and for crude oil, it is 13.93. Hence we can easily reject the hypothesis of a random walk for crude oil, but we cannot reject this hypothesis for lumber, even at the 10 percent level.

Do commodity prices follow random walks? More than a century of data indicates that copper and crude oil prices are not random walks, but the price of lumber is consistent with the random walk hypothesis.[14]

15.4 CO-INTEGRATED TIME SERIES

Regressing one random walk against another can lead to spurious results, in that conventional significance tests will tend to indicate a relationship between the variables when in fact none exists. This is one reason why it is important to test for random walks. If a test fails to reject the hypothesis of a random walk, one can difference the series in question before using it in a regression. Since many economic time series seem to follow random walks, this suggests that one will typically want to difference a variable before using it in a regression. While this is

---

[14] The fact that copper and oil prices do not seem to be random walks does not mean that one can earn an unusually high return by trading these commodities. First, a century is a long time, so even ignoring transaction costs, any excess return from the use of a trading rule is likely to be very small. Second, the mean-reverting behavior that we have found may be due to shifts over time in the risk-adjusted expected return.

acceptable, differencing may result in a loss of information about the long-run relationship between two variables. Are there situations where one can run a regression between two variables in levels, even though both variables are random walks?

There are. Sometimes, two variables will follow random walks, but a *linear combination* of those variables will be stationary. For example, it may be that the variables x_t and y_t are random walks, but the variable $z_t = x_t - \lambda y_t$ is stationary. If this is the case, we say that x_t and y_t are *co-integrated*, and we call λ the *co-integrating parameter*.[15] One can then estimate λ by running an OLS regression of x_t on y_t. (Unlike the case of two random walks that are not co-integrated, here OLS provides a consistent estimator of λ.) Furthermore, the residuals of this regression can then be used to test whether x_t and y_t are indeed co-integrated.

The theory of co-integration, which was developed by Engle and Granger, is important for reasons that go beyond its use as a diagnostic for linear regression.[16] In many cases, economic theory tells us that two variables should be co-integrated, and a test for co-integration is then a test of the theory. For example, although aggregate consumption and disposable income both behave as random walks, we would expect these two variables to move together over the long run, so that a linear combination of the two should be stationary. Another example is the stock market; if stocks are rationally valued, the price of a company's shares should equal the present value of the expected future flow of dividends. This means that although dividends and stock prices both follow random walks, the two series should be co-integrated, with the co-integrating parameter equal to the discount rate used by investors to calculate the present value of earnings.[17]

Suppose one determines, using the Dickey-Fuller test described above, that x_t and y_t are random walks, but Δx_t and Δy_t are stationary. It is then quite easy to test whether x_t and y_t are co-integrated. One simply runs the OLS regression (called the *co-integrating regression*):

$$x_t = \alpha + \beta y_t + \varepsilon_t \qquad (15.40)$$

and then tests whether the residuals, e_t, from this regression are stationary. (If x_t

---

[15] In some situations, x_t and y_t will be vectors of variables, and λ a vector of parameters; λ is then called the *co-integrating vector*. Also, we are assuming that x_t and y_t are both first-order homogeneous nonstationary (also called *integrated of order one*); i.e., the first-differenced series Δx_t and Δy_t are both stationary. More generally, if x_t and y_t are dth-order homogeneous nonstationary (integrated of order d), and $z_t = x_t - \lambda y_t$ is bth-order homogeneous nonstationary, with $b < d$, we say that x_t and y_t are *co-integrated of order d, b*. We will limit our attention to the case of $d = 1$ and $b = 0$.

[16] The theory is set forth in R. F. Engle and C. W. J. Granger, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, vol. 55, pp. 251–276, 1987.

[17] For tests of the present-value model of stock pricing, see J. Y. Campbell and R. J. Shiller, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, vol. 95, pp. 1062–1088, 1987. For other applications, see J. Y. Campbell, "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica*, vol. 55, pp. 1249–1273, 1987, for a study of the co-integration of consumption and income; R. Meese and K. Rogoff, "Was It Real? The Exchange Rate–Interest Differential Relation over the Modern Floating-Rate Period," *Journal of Finance*, vol. 43, pp. 933–947, 1988, for a study of the co-integration of exchange rates and interest differentials.

TABLE 15.3
CRITICAL VALUES FOR TEST
OF DW = 0

Significance level, %	Critical value of DW
1	.511
5	.386
10	.322

and y_t are not co-integrated, any linear combination of them will be nonstationary, and hence the residuals e_t will be nonstationary.) Specifically, we test the hypothesis that e_t is not stationary, i.e., the hypothesis of no co-integration.

A test of the hypothesis that e_t is nonstationary can be done in two ways. First, a Dickey-Fuller test can be performed on the residual series. Alternatively, one can simply look at the Durbin-Watson statistic from the co-integrating regression. Recall from Chapter 6 that the Durbin-Watson statistic is given by

$$DW = \frac{\Sigma(e_t - e_{t-1})^2}{\Sigma(e_t)^2}$$

If e_t is a random walk, the expected value of $(e_t - e_{t-1})$ is zero, so the Durbin-Watson statistic should be close to zero. Thus, one can simply test the hypothesis that DW = 0. For 100 observations, the critical values for this test are shown in Table 15.3.[18] For example, if after running the co-integrating regression we obtain a value of DW of .71, we could reject the hypothesis of no co-integration at the 1 percent level.

Example 15.5 The Co-integration of Consumption and Income

An interesting finding in macroeconomics is that many variables, including aggregate consumption and disposable income, seem to follow random walks. Among other things, this means that the effects of a temporary shock will not tend to dissipate after several years, but instead will be permanent. But even if consumption and disposable income are random walks, the two should tend to move together over long periods, households tend to consume a certain fraction of their disposable income. Thus, over the long term consumption and income should stay in line with each other, i.e., they should be co-integrated.

We will test whether real consumption spending and real disposable income indeed are co-integrated, using quarterly data for the third quarter of

[18] From R. F. Engle and C. W. J. Granger, op. cit., p. 269.

1950 through the first quarter of 1988. We first test whether each variable is a random walk, using the Dickey-Fuller test described in the previous section. For consumption, the unrestricted ESS is 21,203 and the restricted ESS is 22,737; with 151 observations, the F ratio is 5.32. Observe from Table 15.1 that with this value of F we fail to reject the random walk hypothesis, even at the 10 percent level. For disposable income, the unrestricted ESS is 40,418 and the restricted ESS is 42,594, so the F ratio is 3.96. Again we fail to reject the hypothesis of a random walk. (What about first differences of consumption and disposable income? We leave it to the reader to perform a Dickey-Fuller test and show that for first differences we can reject the random walk hypothesis.)

We next run a co-integrating regression of consumption C against disposable income YD.[19] The results are as follows (standard errors in parentheses):

$$C = -133.82 + .9651YD$$
$$(6.109) \quad (.00346)$$

$$R^2 = .9981 \quad\quad s = 23.35 \quad\quad DW = .4936$$

We can use the Durbin-Watson statistic to test whether the residuals from this regression follow a random walk. Comparing the DW of .4936 to the critical values in Table 15.3, we see that we can reject the hypothesis of a random walk at the 5 percent level. The residuals appear to be stationary, so we can conclude that consumption and disposable income are indeed co-integrated.

APPENDIX 15.1 The Autocorrelation Function for a Stationary Process

In this appendix we derive a set of conditions that must hold for an autocorrelation function of a stationary process. Let y_t be a stationary process and let L_t be any linear function of y_t and lags in y_t, for example,

$$L_t = \alpha_1 y_t + \alpha_2 y_{t-1} + \cdots + \alpha_k y_{t-k+1} \tag{A15.1}$$

Now since y_t is stationary, the covariances of y_t are stationary, and

$$\text{Cov}(y_{t+i}, y_{t+j}) = \gamma_{|i-j|} \tag{A15.2}$$

independent of t. Then, by squaring both sides of Eq. (A15.1), we see that the variance of L_t is given by

$$\text{Var}(L_t) = \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} \alpha_i \alpha_j \gamma_{|i-j|} \tag{A15.3}$$

[19] For readers using Citibase data, the corresponding series are GC82 and GYD82.

If the α's are not all 0, the variance of L_t must be greater than 0 and therefore we must have, for all i and j,

$$\gamma_{|i-j|} > 0 \quad\quad \text{for } i = j \tag{A15.4}$$

Now, for n observations, write the covariances of y_t as a matrix:

$$\mathbf{\Gamma}_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} \tag{A15.5}$$

This matrix must be positive definite because the variance of L_t is always greater than zero. Note that

$$\mathbf{\Gamma}_n = \sigma_y^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} = \sigma_y^2 \mathbf{P}_n \tag{A15.6}$$

where $\mathbf{P}_n$ is the matrix of autocorrelations, and is itself positive definite. Thus the determinant of $\mathbf{P}_n$ and its principal minors must be greater than 0.

As an example, let us consider the case of $n = 2$. The condition on the determinant of $\mathbf{P}_n$ becomes

$$\det \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} > 0$$

which implies that

$$1 - \rho_1^2 > 0$$

or

$$-1 < \rho_1 < 1 \tag{A15.7}$$

Similarly, for $n = 3$, it is easy to see that the following three conditions must all hold:

$$-1 < \rho_1 < 1 \tag{A15.8}$$

$$-1 < \rho_2 < 1 \tag{A15.9}$$

$$-1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1 \tag{A15.10}$$

Sets of conditions can also be derived for $n = 4$, $n = 5$, etc., but it should become clear that as the number of observations n becomes large, the number of conditions that must hold also becomes quite large. Although these conditions

can provide an analytical check on the stationarity of a time series, in applied work, it is more typical to judge stationarity from a visual examination of both the series itself and the sample autocorrelation function. For our purposes it will be sufficient to remember that for $k > 0$, $-1 < \rho_k < 1$ for a stationary process.

EXERCISES

15.1 Show that the random walk process with drift is first-order homogeneous nonstationary.

15.2 Consider the time series 1, 2, 3, 4, 5, 6, . . . , 20. Is this series stationary? Calculate the sample autocorrelation function $\hat{\rho}_k$ for $k = 1, 2, \ldots, 5$. Can you explain the shape of this function?

15.3 The data series for the prices of crude oil, copper, and lumber are printed in Table 15.4.

TABLE 15.4
PRICES OF CRUDE OIL, COPPER, AND LUMBER (in 1967 constant dollars)

Obs.	Oil	Copper	Lumber	Obs.	Oil	Copper	Lumber
1870	8.64	41.61	9.13	1929	3.63	36.86	32.65
1871	10.16	47.54	9.70	1930	3.64	29.21	29.84
1872	8.35	70.64	9.75	1931	2.47	21.54	29.39
1873	4.24	61.57	9.98	1932	3.50	16.77	25.42
1874	2.81	54.68	9.93	1933	2.71	20.59	24.97
1875	3.37	53.37	9.45	1934	3.52	21.76	24.97
1876	6.90	49.60	9.60	1935	3.17	20.82	23.51
1877	6.95	51.15	9.74	1936	3.48	22.78	24.34
1878	3.74	47.17	9.75	1937	3.55	29.66	25.12
1879	2.84	49.50	10.43	1938	3.65	24.69	24.59
1880	2.80	52.68	10.09	1939	3.32	27.71	26.55
1881	2.74	46.39	10.90	1940	3.26	27.90	25.38
1882	2.26	49.85	11.11	1941	3.28	26.16	27.47
1883	3.27	42.64	11.05	1942	3.01	23.18	29.98
1884	2.66	41.67	11.79	1943	2.89	22.18	39.74
1885	2.95	35.62	12.02	1944	2.91	22.01	38.76
1886	2.42	32.53	12.32	1945	2.89	21.61	37.78
1887	2.44	31.96	12.44	1946	2.92	22.12	45.31
1888	2.97	52.03	12.03	1947	3.27	27.45	52.84
1889	3.18	45.61	12.03	1948	4.09	26.57	48.76
1890	3.00	52.41	12.28	1949	4.19	24.40	53.94
1891	2.33	43.75	12.19	1950	4.05	25.92	52.30
1892	2.08	41.26	12.60	1951	3.67	26.56	46.18
1893	2.44	38.18	12.55	1952	3.77	27.34	51.13
1894	3.24	36.84	13.72	1953	4.07	32.99	52.86
1895	4.90	41.43	13.07	1954	4.21	33.94	54.91
1896	4.58	44.17	13.67	1955	4.18	42.71	55.48
1897	3.17	45.42	13.17	1956	4.09	46.09	54.20
1898	3.64	46.00	13.32	1957	4.38	31.73	50.88
1899	4.80	63.20	13.68	1958	4.23	27.27	48.78
1900	4.72	55.17	12.59	1959	4.11	32.95	51.24

TABLE 15.4
PRICES OF CRUDE OIL, COPPER, AND LUMBER (in 1967 constant dollars)
(Continued)

Obs.	Oil	Copper	Lumber	Obs.	Oil	Copper	Lumber
1901	4.00	57.19	12.53	1960	4.07	33.83	52.57
1902	3.59	38.16	13.95	1961	4.10	31.64	50.89
1903	4.04	43.00	13.42	1962	4.10	32.28	49.43
1904	3.83	40.91	11.95	1963	4.11	32.38	50.03
1905	2.90	49.03	13.77	1964	4.09	33.79	51.58
1906	2.66	60.50	15.80	1965	3.99	36.23	52.09
1907	2.50	59.52	16.01	1966	3.89	36.27	50.46
1908	2.50	40.74	20.49	1967	3.90	38.20	51.17
1909	2.18	37.36	20.17	1968	3.83	40.78	53.99
1910	1.85	34.99	18.29	1969	3.90	44.60	57.65
1911	2.03	37.01	22.12	1970	3.89	52.26	49.09
1912	2.39	45.79	20.98	1971	4.00	45.13	57.88
1913	3.19	42.50	23.25	1972	3.84	42.49	65.20
1914	2.79	38.75	22.56	1973	3.99	43.73	74.21
1915	2.20	48.19	22.95	1974	5.56	46.81	50.57
1916	3.22	61.68	18.98	1975	5.64	35.11	45.27
1917	3.42	44.88	16.34	1976	5.82	36.22	52.35
1918	3.90	36.34	16.51	1977	5.71	32.48	57.54
1919	3.90	26.15	17.86	1978	5.61	30.23	64.23
1920	5.40	21.96	19.15	1979	6.98	37.77	62.77
1921	4.39	24.85	22.84	1980	10.34	35.99	41.17
1922	4.21	26.85	22.53	1981	13.94	27.22	32.09
1923	3.24	27.75	21.91	1982	12.34	23.37	27.71
1924	3.58	25.69	26.13	1983	11.26	24.80	37.21
1925	4.10	26.22	31.33	1984	10.86	20.76	28.27
1926	4.67	26.74	29.94	1985	10.15	20.47	27.58
1927	3.50	26.22	30.12	1986	5.46	20.91	35.70
1928	3.37	29.26	26.75	1987	6.57	25.61	39.91

(a) Calculate the sample autocorrelation function for each series, and determine whether they are consistent with the Dickey-Fuller test results in Example 15.4. Specifically, do the sample autocorrelation functions for crude oil and copper prices exhibit stationarity? Does the sample autocorrelation function for the price of lumber indicate that the series is nonstationary?

(b) How robust are the Dickey-Fuller test results to the sample size? Divide the sample in half, and for each price series, repeat the Dickey-Fuller tests for each half of the sample.

15.4 Go back to the data for the S&P 500 Common Stock Price Index at the end of the preceding chapter. Would you expect this index to follow a random walk? Perform a Dickey-Fuller test to see whether it indeed does.

15.5 Calculate the sample autocorrelation function for retail auto sales. (Use the data in Table 14.2 at the end of Chapter 14.) Does the sample autocorrelation function indicate seasonality?

LINEAR TIME-SERIES MODELS

We turn now to the main focus of the remainder of this book, the construction of models of stochastic processes and their use in forecasting. Our objective is to develop models that "explain" the movement of a time series by relating it to its own past values and to a weighted sum of current and lagged random disturbances.

While there are many functional forms that can be used, we will use a linear specification. This will allow us to make quantitative statements about the stochastic properties of the models and the forecasts generated by them (e.g., to calculate confidence intervals). In addition, our models apply to stationary processes and to homogeneous nonstationary processes (which can be differenced one or more times to yield stationary processes). Finally, the models are written as equations with fixed estimated coefficients, representing a stochastic structure that does not change over time. (Although models with time-varying coefficients of nonstationary processes have been developed, they are beyond the scope of this book.)

In the first two sections of the chapter we examine simple moving average and autoregressive models for stationary processes. In a moving average model, the process is described completely by a weighted sum of current and lagged random disturbances. In the autoregressive model, the process depends on a weighted sum of its past values and a random disturbance term. In the third section we introduce mixed autoregressive–moving average models. In these models the process is a function of both lagged random disturbances and its past values, as well as a current disturbance term. Even if the original process is nonstationary, it often can be differenced one or more times to produce a new series that is stationary and for which a mixed autoregressive–moving average

model can be constructed. This model can be used to produce a forecast one or more periods into the future, after which the forecasted stationary series can be integrated one or more times to yield a forecast for the original time series. The integrated autoregressive–moving average model provides a general framework for the modeling of homogeneous nonstationary time series.

When building an integrated autoregressive–moving average model for a nonstationary time series, we must first specify how many times the series is to be differenced before a stationary series results. We must also specify the number of autoregressive terms and lagged disturbance terms to be included. We have seen in Chapter 15 that the autocorrelation function can be used to tell us how many times we must difference a homogeneous nonstationary process in order to produce a stationary process. Here we will see how the autocorrelation function can also be used to help determine how many lagged disturbance terms and autoregressive terms should be included in the model.

16.1 MOVING AVERAGE MODELS

In the *moving average process of order q* each observation y_t is generated by a weighted average of random disturbances going back q periods. We denote this process as MA(q) and write its equation as

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \tag{16.1}$$

where the parameters $\theta_1, \ldots, \theta_q$ may be positive or negative.[1]

In the moving average model (and also in the autoregressive model, which will follow) the random disturbances are assumed to be independently distributed across time, i.e., generated by a *white noise* process. In particular, each disturbance term ε_t is assumed to be a normal random variable with mean 0, variance σ_ε^2, and covariance $\gamma_k = 0$ for $k \neq 0$.[2] White noise processes may not occur very commonly, but, as we will see, weighted sums of a white noise process can provide a good representation of processes that are nonwhite.

The reader should observe that the *mean* of the moving average process is independent of time, since $E(y_t) = \mu$. Each ε_t is assumed to be generated by the same white noise process, so that $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, and $E(\varepsilon_t\varepsilon_{t-k}) = 0$ for $k \neq 0$. The process MA (q) is thus described by exactly $q + 2$ parameters, the

---

[1] Following convention, we put a minus sign in front of $\theta_1, \ldots, \theta_q$. In some textbooks, the MA(q) model is written as

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

Be aware of this when reading and interpreting computer output, and as you proceed through the rest of this book.

[2] As we saw in the last chapter, the autocorrelation function for a white noise process is

$$\rho_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$$

mean μ, the disturbance variance σ_ε^2, and the parameters $\theta_1, \theta_2, \ldots, \theta_q$ that determine the weights in the moving average.

Let us now look at the *variance*, denoted by γ_0, of the moving average process of order q:

$$\text{Var}(y_t) = \gamma_0 = E[(y_t - \mu)^2]$$
$$= E(\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 + \cdots + \theta_q^2 \varepsilon_{t-q}^2 - 2\theta_1 \varepsilon_t \varepsilon_{t-1} - \cdots)$$
$$= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 + \cdots + \theta_q^2 \sigma_\varepsilon^2$$
$$= \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \tag{16.2}$$

Note that the expected values of the cross terms are all 0, since we have assumed that the ε_t's are generated by a white noise process for which $\gamma_k = E(\varepsilon_t \varepsilon_{t-k}) = 0$ for $k \neq 0$.

Equation (16.2) imposes a restriction on the values that are permitted for θ_1, $\ldots, \theta_q$. We would expect the variance of y_t to be finite, since otherwise a realization of the random process would involve larger and larger deviations from a fixed reference point as time increased. This, in turn, would violate our assumption of stationarity, since stationarity requires that the probability of being some arbitrary distance from a reference point be invariant with respect to time. Thus, if y_t is the realization of a stationary random process, we must have

$$\sum_{i=1}^{q} \theta_i^2 < \infty \tag{16.3}$$

In a sense this result is trivial, since we have only a finite number of θ_i's, and thus their sum is finite. However, the assumption of a fixed number of θ_i's can be considered to be an approximation to a more general model. A complete model of most random processes would require an infinite number of lagged disturbance terms (and their corresponding weights). Then, as q, the order of the moving average process, becomes infinitely large, we must require that the sum $\sum_{i=0}^{\infty} \theta_i^2$ converge. Convergence will usually occur if the θ's become smaller as i becomes larger. Thus, if we are representing a process, believed to be stationary, by a moving average model of order q, we expect the θ_i's to become smaller as i becomes larger. We will see later that this implies that if the process is stationary, its correlation function ρ_k will become smaller as k becomes larger. This is consistent with our result of the last chapter that one indicator of stationarity is an autocorrelation function that approaches zero.

Now we examine some simple moving average processes, calculating the mean, variance, covariances, and autocorrelation function for each. These statistics are important, first because they provide information that helps characterize the process, and second because they will help us to identify the process when we actually construct models in the next chapter.

We begin with the simplest moving average process, the moving average process of order 1. The process is denoted by MA(1), and its equation is

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{16.4}$$

This process has mean μ and variance $\gamma_0 = \sigma_\varepsilon^2 (1 + \theta_1^2)$. Now let us derive the *covariance* for a one-lag displacement, γ_1:

$$\gamma_1 = E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})]$$
$$= -\theta_1 \sigma_\varepsilon^2 \tag{16.5}$$

In general we can determine the covariance for a k-lag displacement to be

$$\gamma_k = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1})] = 0 \qquad \text{for } k > 1 \tag{16.6}$$

Thus the MA(1) process has a covariance of 0 when the displacement is more than one period. We say, then, that the process has a *memory* of only one period; any value y_t is correlated with y_{t-1} and with y_{t+1}, but with no other time-series values. In effect, the process forgets what happened more than one period in the past. In general the limited memory of a moving average process is important. It suggests that a moving average model provides forecasting information only a limited number of periods into the future.

We can now determine the autocorrelation function for the process MA(1):

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ \\ 0 & k > 1 \end{cases} \tag{16.7}$$

An example of a first-order moving average process might be given by

$$y_t = 2 + \varepsilon_t + .8\varepsilon_{t-1} \tag{16.8}$$

The autocorrelation function for y_t is shown in Fig. 16.1, and a typical realization is shown in Fig. 16.2.

Now let us proceed by examining the moving average process of order 2. The process is denoted by MA(2), and its equation is

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \tag{16.9}$$

This process has mean μ, variance $\sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$, and covariances given by

$$\gamma_1 = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})]$$
$$= -\theta_1 \sigma_\varepsilon^2 + \theta_2 \theta_1 \sigma_\varepsilon^2 = -\theta_1 (1 - \theta_2) \sigma_\varepsilon^2 \tag{16.10}$$
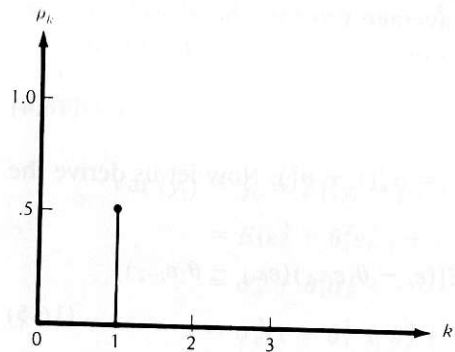
FIGURE 16.1
Autocorrelation function for $y_t = 2 + \varepsilon_t + .8\varepsilon_{t-1}$.

$$\gamma_2 = E[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1\varepsilon_{t-3} - \theta_2\varepsilon_{t-4})]$$

$$= -\theta_2\sigma_\varepsilon^2 \tag{16.11}$$

and

$$\gamma_k = 0 \qquad \text{for } k > 2 \tag{16.12}$$

The autocorrelation function is given by

$$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \tag{16.13}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \tag{16.14}$$

and

$$\rho_k = 0 \qquad \text{for } k > 2 \tag{16.15}$$

The process MA(2) has a memory of exactly two periods, so that the value of y_t is

FIGURE 16.2
Typical realization of $y_t = 2 + \varepsilon_t + .8\varepsilon_{t-1}$.

influenced only by events that took place in the current period, one period back, and two periods back.

An example of a second-order moving average process might be

$$y_t = 2 + \varepsilon_t + .6\varepsilon_{t-1} - .3\varepsilon_{t-2} \tag{16.16}$$

The autocorrelation function is shown in Fig. 16.3, and a typical realization is shown in Fig. 16.4.

We leave to the reader a proof that the moving average process of order q has a memory of exactly q periods, and that its autocorrelation function ρ_k is given by the following (see Exercise 16.3):

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & k = 1, \ldots, q \\ 0 & k > q \end{cases} \tag{16.17}$$

FIGURE 16.3
Autocorrelation function for $y_t = 2 + \varepsilon_t + 6\varepsilon_{t-1} - .3\varepsilon_{t-2}$.

FIGURE 16.4
Typical realization of $y_t = 2 + \varepsilon_t + .6\varepsilon_{t-1} - .3\varepsilon_{t-2}$.
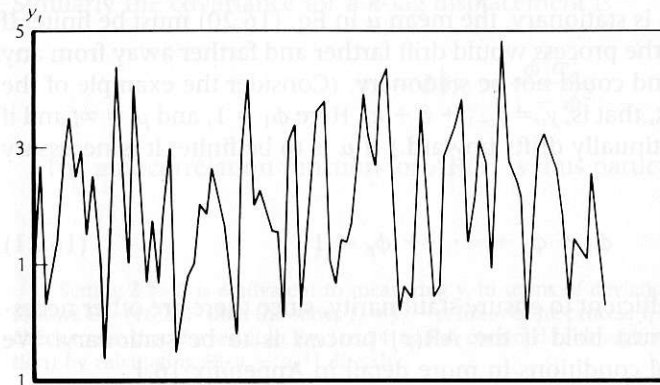
We can now see why the sample autocorrelation function can be useful in specifying the order of a moving average process (assuming that the time series of concern is generated by a moving average process). The autocorrelation function ρ_k for the MA(q) process has q nonzero values and is then 0 for $k > q$. As we proceed through this and later chapters, we will attempt to give the reader an understanding of how the sample autocorrelation function can be used to identify the stochastic process that may have generated a particular time series.

16.2 AUTOREGRESSIVE MODELS

In the *autoregressive process of order p* the current observation y_t is generated by a weighted average of past observations going back p periods, together with a random disturbance in the current period. We denote this process as AR(p) and write its equation as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta + \varepsilon_t \qquad (16.18)$$

Here δ is a constant term which relates (as we will see) to the mean of the stochastic process.

16.2.1 Properties of Autoregressive Models

If the autoregressive process is stationary, then its mean, which we denote by μ, must be invariant with respect to time; that is, $E(y_t) = E(y_{t-1}) = E(y_{t-2}) = \cdots = \mu$. The mean μ is thus given by

$$\mu = \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu + \delta \qquad (16.19)$$

or

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \qquad (16.20)$$

This formula for the mean of the process also gives us a condition for stationarity. If the process is stationary, the mean μ in Eq. (16.20) must be finite. If this were not the case, the process would drift farther and farther away from any fixed reference point and could not be stationary. (Consider the example of the random walk with drift, that is, $y_t = y_{t-1} + \delta + \varepsilon_t$. Here $\phi_1 = 1$, and $\mu = \infty$, and if $\delta > 0$, the process continually drifts upward.) If μ is to be finite, it is necessary that

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1 \qquad (16.21)$$

This condition is not sufficient to ensure stationarity, since there are other necessary conditions that must hold if the AR(p) process is to be stationary. We discuss these additional conditions in more detail in Appendix 16.1.

Now let us examine the properties of some simple autoregressive processes. Again we will determine the mean, covariances, etc., for each. We begin with the first-order process AR(1):

$$y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t \qquad (16.22)$$

This process has mean

$$\mu = \frac{\delta}{1 - \phi_1} \qquad (16.23)$$

and is stationary if $|\phi_1| < 1$. Again, recall that the random walk with drift is a first-order autoregressive process that is *not* stationary. In that process $\phi_1 = 1$, and, as we saw in Chapter 15, the variance of the process becomes larger and larger with time.

Let us now calculate γ_0, the variance of this process about its mean. Assuming stationarity, so that we know that the variance is constant (for $|\phi_1| < 1$), and setting $\delta = 0$ (to scale the process to one that has zero mean), we have[3]

$$\gamma_0 = E[(\phi_1 y_{t-1} + \varepsilon_t)^2] = E(\phi_1^2 y_{t-1}^2 + \varepsilon_t^2 + 2\phi_1 y_{t-1}\varepsilon_t) = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2$$

so that

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \qquad (16.24)$$

We can also calculate the covariances of y_t about its mean:

$$\gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \varepsilon_t)] = \phi_1 \gamma_0 = \frac{\phi_1 \sigma_\varepsilon^2}{1 - \phi_1^2} \qquad (16.25)$$

$$\gamma_2 = E[y_{t-2}(\phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t)] = \phi_1^2 \gamma_0 = \frac{\phi_1^2 \sigma_\varepsilon^2}{1 - \phi_1^2} \qquad (16.26)$$

Similarly the covariance for a k-lag displacement is

$$\gamma_k = \phi_1^k \gamma_0 = \frac{\phi_1^k \sigma_\varepsilon^2}{1 - \phi_1^2} \qquad (16.27)$$

The autocorrelation function for AR(1) is thus particularly simple—it begins

---

[3] Setting $\delta = 0$ is equivalent to measuring y_t in terms of deviations about its mean, since if y follows Eq. (16.22), then the series $\tilde{y}_t = y_t - \mu$ follows the process $\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \varepsilon_t$. The reader can check to see that the result in Eq. (16.24) is also obtained (although with more algebraic manipulation) by calculating $E[(y_t - \mu)^2]$ directly.
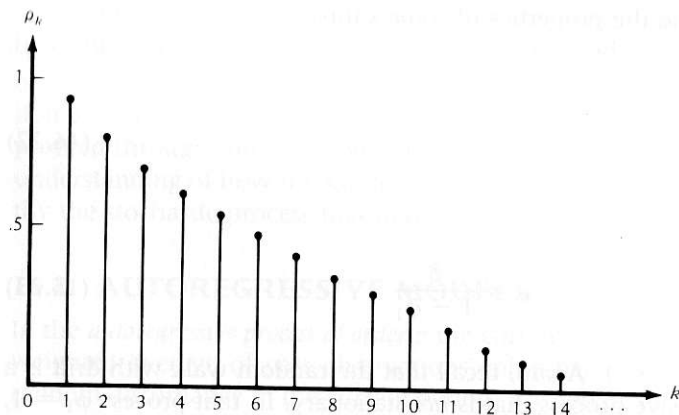
FIGURE 16.5
Autocorrelation function for $y_t = .9y_{t-1} + 2 + \varepsilon_t$.

at $\rho_0 = 1$ and then declines geometrically:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k \qquad (16.28)$$

Note that this process has an *infinite memory*. The current value of the process depends on *all past values*, although the magnitude of this dependence declines with time.[4]

An example of a first-order autoregressive process would be the process defined by

$$y_t = .9y_{t-1} + 2 + \varepsilon_t \qquad (16.29)$$

The autocorrelation function for this process is shown in Fig. 16.5, and a typical realization is shown in Fig. 16.6. The realization differs from that of a first-order moving average process in that each observation is highly correlated with those surrounding it, resulting in discernible overall up-and-down patterns.

Let us now look at the second-order autoregressive process AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta + \varepsilon_t \qquad (16.30)$$

[4] It can be shown that if the AR(1) process is stationary, it is equivalent to a moving average process of *infinite order* (and thus with infinite memory). In fact, for any stationary autoregressive process of any order there exists an equivalent moving average process of infinite order (so that the autoregressive process is *invertible* into a moving average process). Similarly, if certain *invertibility conditions* are met (and these will be discussed in Appendix 16.1), any finite-order moving average process has an equivalent autoregressive process of infinite order. For a more detailed discussion of invertibility, the reader is referred to G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (San Francisco: Holden-Day, 1970); C. Nelson, *Applied Time Series Analysis* (San Francisco: Holden-Day, 1973), chap. 3; or C. W. J. Granger and P. Newbold, *Forecasting Economic Time Series* (New York: Academic, 1986).
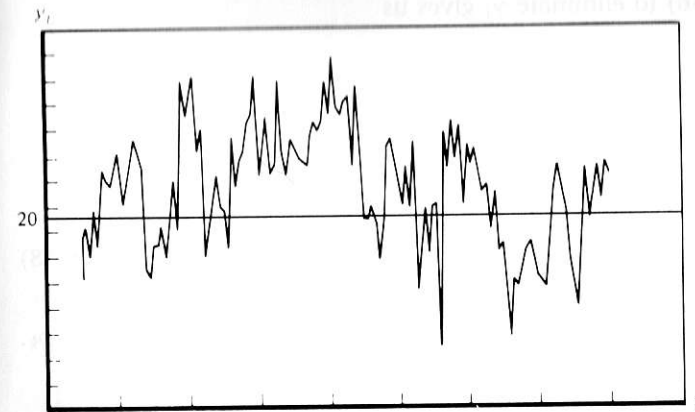
FIGURE 16.6
Typical realization of the process $y_t = .9y_{t-1} + 2 + \varepsilon_t$.

The process has mean

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2} \qquad (16.31)$$

and a necessary condition for stationarity is that $\phi_1 + \phi_2 < 1$.[5]

Let us now calculate the variances and covariances of y_t (when y_t is measured in deviations form):

$$\gamma_0 = E[y_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2 \qquad (16.32)$$

$$\gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1\gamma_0 + \phi_2\gamma_1 \qquad (16.33)$$

$$\gamma_2 = E[y_{t-2}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1\gamma_1 + \phi_2\gamma_0 \qquad (16.34)$$

and in general, for $k \geq 2$,

$$\gamma_k = E[y_{t-k}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} \qquad (16.35)$$

We can solve Eqs. (16.32), (16.33), and (16.34) simultaneously to get γ_0 in terms of ϕ_1, ϕ_2, and σ_ε^2. Equation (16.33) can be rewritten as

$$\gamma_1 = \frac{\phi_1\gamma_0}{1 - \phi_2} \qquad (16.36)$$

Substituting Eq. (16.34) into Eq. (16.32) yields

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\phi_1\gamma_1 + \phi_2^2\gamma_0 + \sigma_\varepsilon^2 \qquad (16.37)$$

[5] Necessary and sufficient conditions are presented in Appendix 16.1.

Now using Eq. (16.36) to eliminate γ_1 gives us

$$\gamma_0 = \frac{\phi_1^2 \gamma_0}{1 - \phi_2} + \frac{\phi_2 \phi_1^2 \gamma_0}{1 - \phi_2} + \phi_2^2 \gamma_0 + \sigma_\varepsilon^2$$

which, after rearranging, yields

$$\gamma_0 = \frac{(1 - \phi_2)\sigma_\varepsilon^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \tag{16.38}$$

These equations can also be used to derive the autocorrelation function ρ_k. From Eqs. (16.34) and (16.36),

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \tag{16.39}$$

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \tag{16.40}$$

From Eq. (16.35) one can see that for $k \geq 2$,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \tag{16.41}$$

and this can be used to calculate the autocorrelation function for $k > 2$.

A comment is in order regarding Eqs. (16.39) and (16.40), which are called the *Yule-Walker equations*. Suppose we have the sample autocorrelation function for a time series which we believe was generated by a second-order autoregressive process. We could then measure ρ_1 and ρ_2 and substitute these numbers into Eqs. (16.39) and (16.40). We would then have two algebraic equations which could be solved simultaneously for the two unknowns ϕ_1 and ϕ_2. Thus, we could use the Yule-Walker equations to obtain estimates of the autoregressive parameters ϕ_1 and ϕ_2.

Let us look at an example of a second-order autoregressive process:

$$y_t = .9y_{t-1} - .7y_{t-2} + 2 + \varepsilon_t \tag{16.42}$$

The autocorrelation function for this process is shown in Fig. 16.7. Note that it is a sinusoidal function that is geometrically damped. As we will see from further examples, autocorrelation functions for autoregressive processes (of order greater than 1) are typically geometrically damped, oscillating, sinusoidal functions.

The reader should note that *realizations* of second- (and higher-) order autoregressive processes may or may not be cyclical, depending on the numerical values of the parameters ϕ_1, ϕ_2, etc. Equation (16.30), for example, is a second-
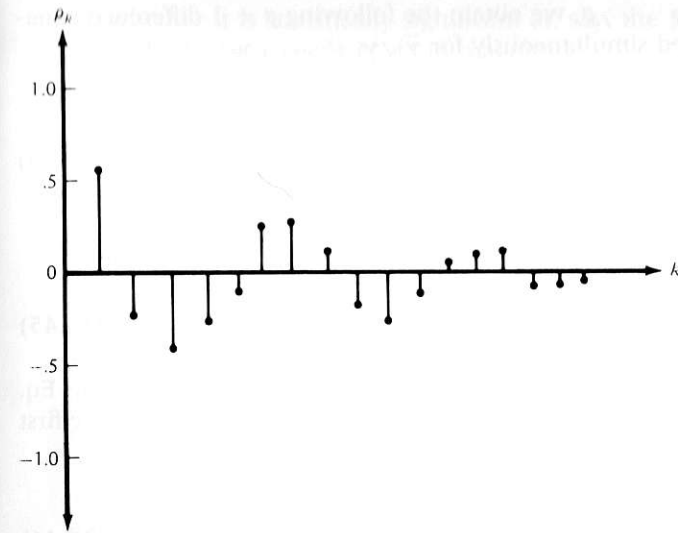
FIGURE 16.7
Autocorrelation function for $y_t = .9y_{t-1} - .7y_{t-2} + 2 + \varepsilon_t$.

order difference equation in y_t (with an additive error term). We saw in Chapter 13 that the values of ϕ_1 and ϕ_2 determine whether the solution to this difference equation is oscillatory.

16.2.2 The Partial Autocorrelation Function

One problem in constructing autoregressive models is identifying the *order* of the underlying process. For moving average models this is less of a problem, since if the process is of order q the sample autocorrelations should all be close to zero for lags greater than q. (Bartlett's formula provides approximate standard errors for the autocorrelations, so that the order of a moving average process can be determined from significance tests on the sample autocorrelations.) Although some information about the order of an autoregressive process can be obtained from the oscillatory behavior of the sample autocorrelation function, much more information can be obtained from the *partial autocorrelation function*.

To understand what the partial autocorrelation function is and how it can be used, let us first consider the covariances and autocorrelation function for the autoregressive process of order p. First, notice that the covariance with displacement k is determined from

$$\gamma_k = E[y_{t-k}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t)] \tag{16.43}$$

Now letting $k = 0, 1, \ldots, p$, we obtain the following $p + 1$ difference equations that can be solved simultaneously for $\gamma_0, \gamma_1, \ldots, \gamma_p$:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_\varepsilon^2$$
$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1 + \cdots + \phi_p\gamma_{p-1}$$
$$\cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \qquad (16.44)$$
$$\gamma_p = \phi_1\gamma_{p-1} + \phi_2\gamma_{p-2} + \cdots + \phi_p\gamma_0$$

For displacements k greater than p the covariances are determined from

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} \qquad (16.45)$$

Now by dividing the left-hand and right-hand sides of the equations in Eq. (16.44) by γ_0, we can derive a set of p equations that together determine the first p values of the autocorrelation function:

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1}$$
$$\cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \qquad (16.46)$$
$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p$$

For displacement k greater than p we have, from Eq. (16.45),

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad (16.47)$$

The equations in Eq. (16.46) are the Yule-Walker equations; if $\rho_1, \rho_2, \ldots, \rho_p$ are known, then the equations can be solved for $\phi_1, \phi_2, \ldots, \phi_p$.

Unfortunately, solution of the Yule-Walker equations as presented in Eq. (16.46) requires knowledge of p, the order of the autoregressive process. Therefore, we solve the Yule-Walker equations for *successive values of p*. In other words, suppose we begin by hypothesizing that $p = 1$. Then Eqs. (16.46) boil down to $\rho_1 = \phi_1$ or, using the sample autocorrelations, $\hat{\rho}_1 = \hat{\phi}_1$. Thus, if the calculated value $\hat{\phi}_1$ is significantly different from zero, we know that the autoregressive process is *at least* order 1. Let us denote this value $\hat{\phi}_1$ by a_1.

Now let us consider the hypothesis that $p = 2$. To do this we just solve the Yule-Walker equations [Eq. (16.46)] for $p = 2$. Doing this gives us a new set of estimates $\hat{\phi}_1$ and $\hat{\phi}_2$. If ϕ_2 is significantly different from zero we can conclude that the process is *at least* order 2, while if $\hat{\phi}_2$ is approximately zero, we can conclude that $p = 1$. Let us denote the value $\hat{\phi}_2$ by a_2.

We now repeat this process for successive values of p. For $p = 3$ we obtain an estimate of $\hat{\phi}_3$ which we denote by a_3, for $p = 4$ we obtain $\hat{\phi}_4$ which we denote by a_4, etc. We call this series $a_1, a_2, a_3, \ldots$ the *partial autocorrelation function* and note that we can infer the order of the autoregressive process from its behavior. In particular, if the true order of the process is p, we should observe that $a_j \approx 0$ for $j > p$.

To test whether a particular a_j is zero, we can use the fact that it is approximately normally distributed, with mean zero and variance $1/T$. Hence we can check whether it is statistically significant at, say, the 5 percent level by determining whether it exceeds $2/\sqrt{T}$ in magnitude.

Example 16.1 Inventory Investment In Chapter 15 we examined the behavior of real nonfarm inventory investment over the period 1952 through the second quarter of 1988. (The series itself is shown in Fig. 15.3, and its sample autocorrelation function is shown in Fig. 15.4.) We concluded that the series is stationary because the sample autocorrelation function falls toward zero quickly as the number of lags increases.

Figure 16.8 shows the partial autocorrelation function for our inventory investment series. Observe that the partial autocorrelations become close to zero after about four lags. Since there are 146 data points in the sample, a partial autocorrelation is statistically significant at the 5 percent level only if it is larger in magnitude than $2/\sqrt{146} = .166$. There are no partial autocorrelations beyond four lags that are this large. We can conclude from this that to the extent that inventory investment follows an autoregressive process, the order of that process should not be greater than 4. We will take this information into account when we construct a time-series model for inventory investment in Chapter 19.

FIGURE 16.8
Inventory investment: partial autocorrelation function.

16.3 MIXED AUTOREGRESSIVE–MOVING AVERAGE MODELS

Many stationary random processes cannot be modeled as purely moving average or as purely autoregressive, since they have the qualities of both types of processes. The logical extension of the models presented in the last two sections is the *mixed autoregressive–moving average process of order* (p, q). We denote this process as ARMA(p, q) and represent it by

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (16.48)$$

We assume that the process is stationary, so that its mean is constant over time and is given by

$$\mu = \phi_1 \mu + \cdots + \phi_p \mu + \delta$$

or

$$\mu = \frac{\delta}{1 - \phi_1 - \cdots - \phi_p} \quad (16.49)$$

This gives a necessary condition for the stationarity of the process, that is,

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1 \quad (16.50)$$

Now let us consider the simplest mixed autoregressive–moving average process, the process ARMA(1, 1):

$$y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (16.51)$$

The variances and covariances of this process are determined jointly as follows (setting $\delta = 0$):

$$\gamma_0 = E[y_t(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = E[(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})^2]$$
$$= \phi_1^2 \gamma_0 - 2\phi_1 \theta_1 E[y_{t-1}\varepsilon_{t-1}] + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \quad (16.52)$$

Since $E(y_{t-1}\varepsilon_{t-1}) = \sigma_\varepsilon^2$, we have

$$\gamma_0(1 - \phi_1^2) = \sigma_\varepsilon^2(1 + \theta_1^2 - 2\phi_1\theta_1) \quad (16.53)$$

so that the variance is given by[6]

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \sigma_\varepsilon^2 \quad (16.54)$$

[6] For $|\phi_1| < 1$.

We can now determine the covariances $\gamma_1, \gamma_2, \ldots$, recursively:

$$\gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1})] = \phi_1\gamma_0 - \theta_1\sigma_\varepsilon^2$$
$$= \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2}\sigma_\varepsilon^2 \quad (16.55)$$

$$\gamma_2 = E[y_{t-2}(\phi_1 y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1})] = \phi_1\gamma_1 \quad (16.56)$$

and similarly,

$$\gamma_k = \phi_1\gamma_{k-1} \quad k \geq 2 \quad (16.57)$$

The autocorrelation function, then, is given by

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \quad (16.58)$$

and for displacement k greater than 1,

$$\rho_k = \phi_1\rho_{k-1} \quad k \geq 2 \quad (16.59)$$

Thus, the autocorrelation function begins at its starting value ρ_1 (which is a function of both ϕ_1 and θ_1) and then decays geometrically from this starting value. This reflects the fact that the moving average part of the process has a memory of only one period.

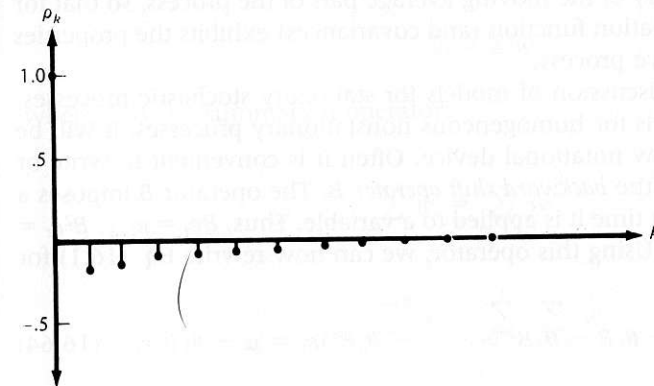Let us examine the autocorrelation functions of some typical ARMA(1, 1) processes. The autocorrelation function for the process
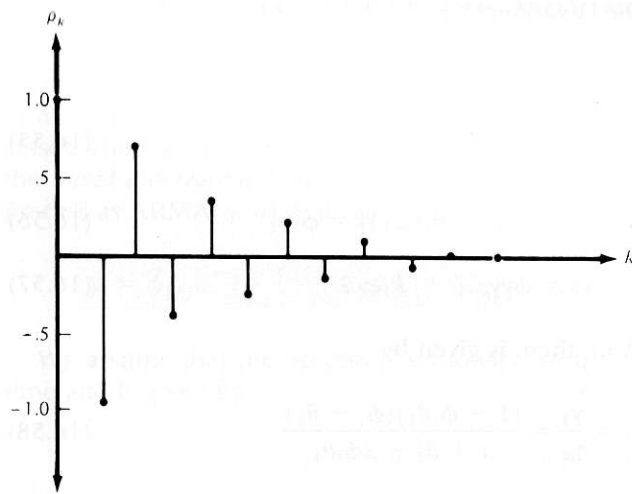
$$y_t = .8y_{t-1} + 2 + \varepsilon_t - .9\varepsilon_{t-1} \quad (16.60)$$

is shown in Fig. 16.9. The starting value ρ_1 is negative, and the function decays toward 0 from this value.

FIGURE 16.9
Autocorrelation function for $y_t = .8y_{t-1} + 2 + \varepsilon_t - .9\varepsilon_{t-1}$.

FIGURE 16.10
Autocorrelation function for $y_t = -.8y_{t-1} + 2 + \varepsilon_t + .9\varepsilon_{t-1}$.

The autocorrelation function for the process

$$y_t = -.8y_{t-1} + 2 + \varepsilon_t + .9\varepsilon_{t-1} \qquad (16.61)$$

will exhibit oscillatory behavior, as shown in Fig. 16.10. Note that it oscillates between positive and negative values, since ϕ_1 is negative.

For higher-order processes, i.e., the general ARMA(p, q) process, the variance, covariances, and autocorrelation function are solutions to difference equations that usually cannot be solved by inspection. It can be shown easily, however, that

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} \qquad k \geq q + 1 \qquad (16.62)$$

and thus

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad k \geq q + 1 \qquad (16.63)$$

Note that q is the memory of the moving average part of the process, so that for $k \geq q + 1$ the autocorrelation function (and covariances) exhibits the properties of a purely autoregressive process.

This completes our discussion of models for stationary stochastic processes. Before we turn to models for homogeneous nonstationary processes, it will be useful to introduce a new notational device. Often it is convenient to write or describe time lags using the *backward shift operator B*. The operator B imposes a one-period time lag each time it is applied to a variable. Thus, $B\varepsilon_t = \varepsilon_{t-1}$, $B^2\varepsilon_t = \varepsilon_{t-2}, \ldots, B^n\varepsilon_t = \varepsilon_{t-n}$. Using this operator, we can now rewrite Eq. (16.1) for the MA(q) process as

$$y_t = \mu + (1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q)\varepsilon_t = \mu + \theta(B)\varepsilon_t \qquad (16.64)$$

where $\theta(B)$ denotes a polynomial function of the operator B. Similarly Eq. (16.18) for the AR(p) process can be rewritten as

$$(1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p)y_t = \delta + \varepsilon_t \qquad (16.65)$$

or

$$\phi(B)y_t = \delta + \varepsilon_t \qquad (16.66)$$

Finally, Eq. (16.48) for the ARMA(p, q) process can be rewritten as

$$(1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p)y_t = \delta + (1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q)\varepsilon_t \qquad (16.67)$$

or

$$\phi(B)y_t = \delta + \theta(B)\varepsilon_t \qquad (16.68)$$

16.4 HOMOGENEOUS NONSTATIONARY PROCESSES: ARIMA MODELS

In practice, many of the time series we will work with are nonstationary, so that the characteristics of the underlying stochastic process change over time. In this section we construct models for those nonstationary series which can be transformed into stationary series by differencing one or more times. We say that y_t is *homogeneous nonstationary of order d* if

$$w_t = \Delta^d y_t \qquad (16.69)$$

is a stationary series. Here Δ denotes differencing, i.e.,

$$\Delta y_t = y_t - y_{t-1} \qquad \Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

and so forth. A discussion of the autoregressive characteristics of homogeneous nonstationary series is given in Appendix 16.1.

Observe that if we have a series w_t, we can get back to y_t by *summing* w_t a total of d times. We write this as

$$y_t = \Sigma^d w_t \qquad (16.70)$$

where Σ is the summation operator:

$$\sum w_t = \sum_{i=-\infty}^{t} w_i \qquad (16.71)$$

$$\sum{}^2 w_t = \sum_{j=-\infty}^{t} \sum_{i=-\infty}^{j} w_i \qquad (16.72)$$

and so forth. Note that the summation operator Σ is just the *inverse* of the difference operator Δ. Since $\Delta y_t = y_t - y_{t-1}$, we can write that $\Delta = 1 - B$, and thus $\Sigma = \Delta^{-1} = (1 - B)^{-1}$.

When computing this sum for an actual time series, we begin with the first observation on the original undifferenced series (y_0) and then add successive values of the differenced series. Thus if $w_t = \Delta y_t$, we would compute y_t from

$$y_t = \sum w_t = \sum_{i=-\infty}^{t} w_i = \sum_{i=-\infty}^{0} w_i + \sum_{i=1}^{t} w_i = y_0 + w_1 + w_2 + \cdots + w_t$$

$$(16.73)$$

If y_t had been differenced twice, so that $w_t = \Delta^2 y_t$, we could compute y_t from w_t by summing w_t twice.[7]

After we have differenced the series y_t to produce the stationary series w_t, we can model w_t as an ARMA process. *If $w_t = \Delta^d y_t$, and w_t is an ARMA(p, q) process, then we say that y_t is an integrated autoregressive–moving average process of order (p, d, q), or simply* ARIMA(p, d, q). We can write the equation for the process ARIMA(p, d, q), using the backward shift operator, as

$$\phi(B)\Delta^d y_t = \delta + \theta(B)\varepsilon_t \qquad (16.74)$$

with
$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (16.75)$$

and
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \qquad (16.76)$$

We call $\phi(B)$ the *autoregressive operator* and $\theta(B)$ the *moving average operator*.

Note that the mean of $w_t = \Delta^d y_t$ is given by

$$\mu_w = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \qquad (16.77)$$

Thus if δ is not equal to 0, the *integrated series* y_t will have a built-in *deterministic trend*. Suppose, for example, that $d = 1$ and $\delta > 0$. Then $y_t = \Sigma w_t$ will grow linearly over time. An example of such a series might be the one drawn in Fig. 16.11. The series has a linear time trend that is independent of the random disturbances, i.e., that is deterministic. The series drawn in Fig. 16.12, on the

---

[7] Summing w_t the first time gives us Δy_t:

$$\Delta y_t = \sum w_t = \sum_{i=-\infty}^{0} w_i + \sum_{i=1}^{t} w_i = \Delta y_0 + w_1 + w_2 + \cdots + w_t$$

Now summing Δy_t yields y_t:

$$y_t = \Sigma(\Delta y_t) = \Sigma(\Delta y_0 + w_1 + w_2 + \cdots + w_t) = y_0 + (\Delta y_0 + w_1) + (\Delta y_0 + w_1 + w_2) + \cdots$$
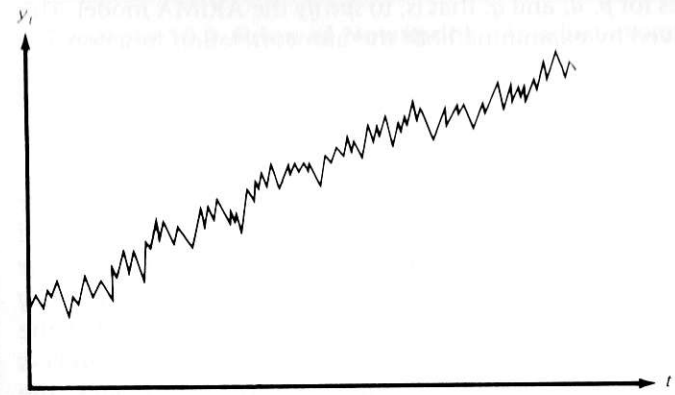$$+ (\Delta y_0 + w_1 + w_2 + \cdots + w_t)$$

FIGURE 16.11
An ARIMA process with $d = 1$.

other hand, has an average slope that is increasing linearly in time. This series might have been generated by a process that is ARIMA with $d = 2$ and $\delta > 0$. Thus $w_t = \Delta^2 y_t$ will have no time trend, $\Sigma w_t = \Delta y_t$ will have a linear time trend, and $\Sigma\Sigma w_t = y_t$ will have a time trend whose rate of increase is constant.

It is possible that the stationary series w_t will not be mixed, i.e., will be completely autoregressive or moving average. If w_t is just AR(p), then we call y_t an integrated autoregressive process of order (p, d), and denote it as ARI(p, d, 0). If w_t is just MA(q), then we call y_t an integrated moving average process of order (d, q), and denote it as IMA(0, d, q).

16.5 SPECIFICATION OF ARIMA MODELS

We have seen that any homogeneous nonstationary time series can be modeled as an ARIMA process of order (p, d, q). The practical problem is to choose the

FIGURE 16.12
An ARIMA process with $d = 2$.
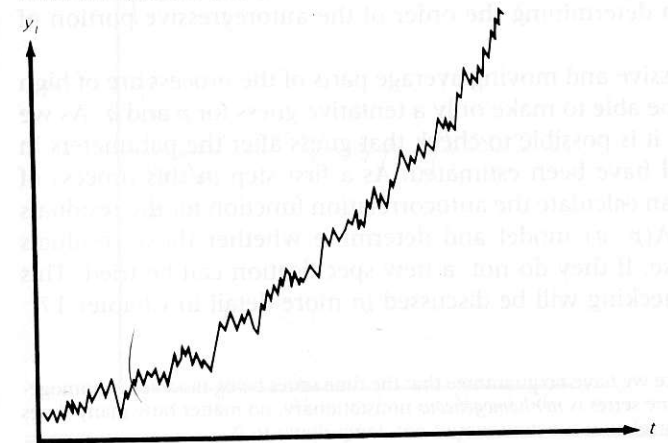
most appropriate values for p, d, and q, that is, to *specify* the ARIMA model. This problem is partly resolved by examining both the autocorrelation function and the partial autocorrelation function for the time series of concern.

Given a series y_t that one would like to model, the first problem is to determine the degree of homogeneity d, that is, the number of times that the series must be differenced to produce a stationary series. To do this, we make use of the fact that *the autocorrelation function ρ_k for a stationary series must approach 0 as the displacement k becomes large*. To see why, consider a stationary ARMA process of order (p, q). We know that the autocorrelation function for the *moving average part* of this process becomes 0 for $k > q$, as the process has a memory of only q periods. Thus, if y_t is MA(q), then $\rho_k = 0$ for $k > q$. We also know that the autocorrelation function for the *autoregressive part* of a stationary ARMA process is geometrically damped (see the examples in Figs. 16.5 to 16.7). Finally, the autocorrelation function for the complete ARMA process has moving average characteristics for the first $q - p$ periods, but after that it is autoregressive in character; i.e., it has an envelope that declines geometrically.

To specify d, first examine the autocorrelation function of the original series y_t and determine whether it is stationary. If it is not, difference the series and examine the autocorrelation function for Δy_t. Repeat this process until a value for d is reached such that $\Delta^d y_t$ is stationary; i.e., the autocorrelation function goes to 0 as k becomes large.[8] One should also examine the time series itself; if it appears to have an overall trend, it is probably not stationary.

After d is determined, one can work with the stationary series $w_t = \Delta^d y_t$ and examine both its autocorrelation function and its partial autocorrelation function to determine possible specifications for p and q. For low-order processes this is not too difficult, since the autocorrelation functions for processes such as AR(1), AR(2), MA(1), MA(2), and ARMA(1, 1) are easy to recognize and distinguish from each other (see Figs. 16.1 to 16.10). However, if the time series cannot be modeled as a low-order ARMA process, the specification of p and q becomes more difficult and requires close inspection of the full and partial autocorrelation functions. For example, spikes in the autocorrelation function are indicative of moving average terms, and the partial autocorrelation function can be used for guidance in determining the order of the autoregressive portion of the process.

If both the autoregressive and moving average parts of the process are of high order, one may at best be able to make only a tentative guess for p and q. As we will see later, however, it is possible to check that guess after the parameters in the ARMA(p, q) model have been estimated. As a first step in this process of *diagnostic checking* one can calculate the autocorrelation function for the residuals of the estimated ARMA(p, q) model and determine whether those residuals appear to be white noise. If they do not, a new specification can be tried. This process of diagnostic checking will be discussed in more detail in Chapter 17.

[8] Remember that in practice we have no guarantee that the time series being modeled is homogeneous nonstationary. If the time series is *nonhomogeneous* nonstationary, no matter how many times it was differenced, the autocorrelation function would not damp down to 0.

Example 16.2 Price of Newsprint As a first example of model specification, let us examine a quarterly series for the average price of newsprint in the United States over the period 1965-2 through 1977-3 (50 data points). The series itself (not shown here) rises steadily over time, indicating that it is nonstationary. However, the differenced series Δy_t does appear to be stationary, as can be seen from its sample autocorrelation function in Fig. 16.13. The autocorrelation function has the damped sinusoidal shape of a second-order autoregressive process and no spikes indicative of moving average terms. The partial autocorrelation function, shown in Fig. 16.14, has significant spikes at lags 1 and 2, confirming a second-order autoregressive interpretation of the differenced series. We might thus estimate an ARI(2, 1, 0) model.
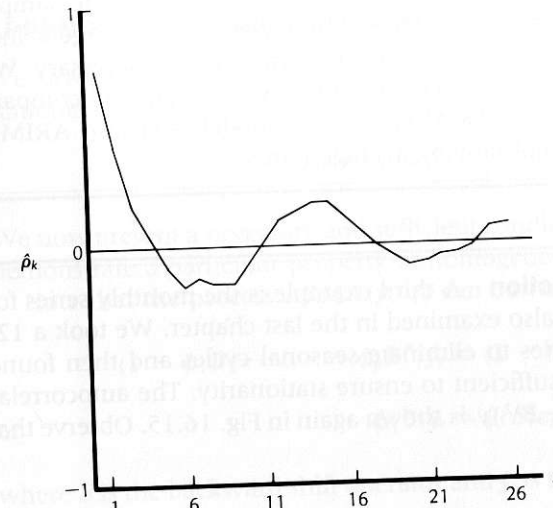
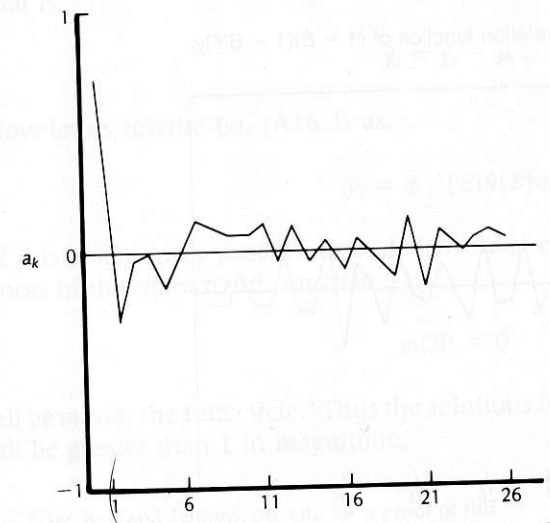FIGURE 16.13
Newsprint price: autocorrelation function of Δy_t.

FIGURE 16.14
Newsprint price: partial autocorrelation function of Δy_t.

Example 16.3 Interest Rates As a second example of model specification, go back to the series for the 3-month Treasury bill rate that we examined in Chapter 15. After differencing the series and examining the sample autocorrelation functions, we established that it was probably first-order homogeneous nonstationary, so that d equals 1 in a specification of an ARIMA model. Now if we examine the autocorrelation function for Δy_t in Fig. 15.10 in more detail, we see that it exhibits moving average properties that are first or second order; i.e., it begins decaying after the point $k = 2$.

What about the autoregressive properties of the interest rate series? For $k \geq 1$ none of the sample autocorrelations exceed .25 in magnitude, suggesting that only a few autoregressive terms might suffice. Hence, one could begin by estimating an ARIMA(2, 1, 2) model. On the other hand, the sample autocorrelations remain significantly different from zero even for large values of k, suggesting that many more autoregressive terms may be necessary. We explore this possibility in the next chapter, where we estimate and compare an ARIMA(2, 1, 2) model, an ARIMA(12, 1, 2) model, and also ARIMA models that include additional moving average terms.

Example 16.4 Hog Production A third example is the monthly series for hog production which was also examined in the last chapter. We took a 12-month difference of the series to eliminate seasonal cycles and then found that differencing once was sufficient to ensure stationarity. The autocorrelation function for $(1 - B)(1 - B^{12})y_t$ is shown again in Fig. 16.15. Observe that

FIGURE 16.15
Monthly hog production: autocorrelation function of $(1 - B)(1 - B^{12})y_t$.
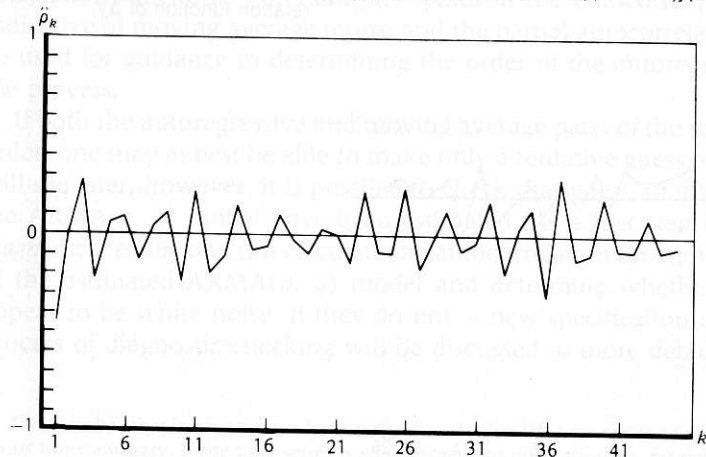
the sample autocorrelation function begins declining immediately at $k = 1$ and has peaks roughly once every three periods. We might thus suspect that $(1 - B)(1 - B^{12})y_t$ is autoregressive of order 3, so that y_t could be specified by the model:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)(1 - B^{12})y_t = \varepsilon_t \quad (16.78)$$

Readers should not be disturbed at this point if they find this process of model specification somewhat bewildering. We will go through several more examples in Chapter 19.

APPENDIX 16.1 Stationarity, Invertibility, and Homogeneity

We saw before that a necessary condition for an ARMA(p, q) process to be stationary is that

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1 \quad (A16.1)$$

We now present a necessary and sufficient condition for stationarity and use it to demonstrate a particular property of homogeneous nonstationary processes.

Note that the process ARMA(p, q) can be written in the form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)\bar{y}_t = (1 - \theta_1 B - \cdots - \theta_q B^q)\varepsilon_t \quad (A16.2)$$

or

$$\phi(B)\bar{y}_t = \theta(B)\varepsilon_t \quad (A16.3)$$

where B is the backward shift operator and $\bar{y}_t$ is the deviation of y_t from its mean, that is,

$$\bar{y}_t = y_t - \mu \quad (A16.4)$$

Now let us rewrite Eq. (A16.3) as

$$\bar{y}_t = \phi^{-1}(B)\theta(B)\varepsilon_t \quad (A16.5)$$

If y_t is a stationary process, then $\phi^{-1}(B)$ must converge. This requires that the roots of the *characteristic equation*

$$\phi(B) = 0 \quad (A16.6)$$

all be *outside* the unit circle.[9] Thus the solutions $B_1, \ldots, B_p$ to Eq. (A16.6) must all be greater than 1 in magnitude.

[9] See Box and Jenkins, op. cit., for a proof of this.

Now suppose that the process $\tilde{y}_t$ in Eq. (A16.3) is *nonstationary* but in such a way that exactly d of the roots of $\phi(B)$ are on the unit circle and the remainder are outside the unit circle. This process can then be rewritten in the form

$$\omega(B)(1 - B)^d \tilde{y}_t = \theta(B)\varepsilon_t \qquad (A16.7)$$

where $\omega(B)$ is a stationary autoregressive operator of order $p - d$ and the operator $(1 - B)^d$ has d roots all equal to unity. But $1 - B$ is a *first-difference* operator, that is,

$$(1 - B)\tilde{y}_t = \Delta\tilde{y}_t = \tilde{y}_t - \tilde{y}_{t-1} \qquad (A16.8)$$

Thus Eq. (A16.7) can be rewritten as

$$\omega(B)\Delta^d\tilde{y}_t = \theta(B)\varepsilon_t \qquad (A16.9)$$

or

$$\omega(B)w_t = \theta(B)\varepsilon_t \qquad (A16.10)$$

where $w_t = \Delta^d\tilde{y}_t$ is stationary, since it resulted from differencing $\tilde{y}_t$ d times. We call $\tilde{y}_t$ *homogeneous nonstationary with order d,* and we note the conclusion that such a process has an autoregressive operator $\phi(B)$ such that

$$\phi(B) = \omega(B)(1 - B)^d \qquad (A16.11)$$

where the roots of $\omega(B)$ are all outside the unit circle.

Analogous to the stationarity condition for the autoregressive operator is the *invertibility condition* for the moving average operator. We say that y_t is invertible if we can write Eq. (A16.3) as

$$\theta^{-1}(B)\phi(B)\tilde{y}_t = \varepsilon_t \qquad (A16.12)$$

i.e., if the moving average part of the ARMA process can be inverted into a purely autoregressive process. Now, if y_t is invertible, $\theta^{-1}(B)$ must converge. This requires that the roots of the *characteristic equation*

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0 \qquad (A16.13)$$

must all lie outside the unit circle; i.e., the solutions $B_1, B_2, \ldots, B_q$ to Eq. (A16.13) must all be greater than 1 in absolute value.[10]

---

[10] This invertibility condition is proved in detail in U. Grenander and M. Rosenblatt, *Statistical Analyses of Stationary Time Series* (New York: Wiley, 1957). A brief discussion is also presented in Box and Jenkins, op. cit.

As an example, consider the first-order moving average process (that is, $q = 1$), whose characteristic equation is

$$1 - \theta_1 B = 0 \qquad (A16.14)$$

Then the invertibility condition becomes

$$|B| = \frac{1}{|\theta_1|} > 1 \qquad (A16.15)$$

or

$$|\theta_1| < 1$$

For the second-order moving average process ($q = 2$) the characteristic equation is

$$1 - \theta_1 B - \theta_2 B^2 = 0 \qquad (A16.16)$$

and

$$B = \frac{-\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{2\theta_2} \qquad (A16.17)$$

Both these values of B must be outside the unit circle, which implies that

$$\theta_2 + \theta_1 < 1 \qquad (A16.18)$$

$$\theta_2 - \theta_1 < 1 \qquad (A16.19)$$

and

$$|\theta_2| < 1 \qquad (A16.20)$$

EXERCISES

16.1 Calculate the covariances γ_k for MA(3), the moving average process of order 3. Determine the autocorrelation function for this process. Plot the autocorrelation function for the MA(3) process

$$y_t = 1 + \varepsilon_t + .8\varepsilon_{t-1} - .5\varepsilon_{t-2} + .3\varepsilon_{t-3}$$

16.2 What are the characteristics that one would expect of a realization of the following MA(1) process?

$$y_t = 1 + \varepsilon_t + .8\varepsilon_{t-1}$$

How would these characteristics differ from those of a realization of the following MA(1) process?

$$y_t = 1 + \varepsilon_t - .8\varepsilon_{t-1}$$

16.3 Show that the covariances γ_k of MA(q), the moving average process of order q, are given by

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2 & k = 1, \ldots, q \\ 0 & k > q \end{cases}$$

and that the autocorrelation function for MA(q) is given by

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2} & k = 1, \ldots, q \\ 0 & k > q \end{cases}$$

as in Eq. (16.17).

16.4 Derive the autocorrelation function for the ARMA(2, 1) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

that is, determine ρ_1, ρ_2, etc., in terms of ϕ_1, ϕ_2, and θ_1. Draw this autocorrelation function for $\phi_1 = .6$, $\phi_2 = .3$, and $\theta_1 = .9$. Repeat for $\phi_1 = .6$, $\phi_2 = .3$, and $\theta_1 = -.9$. Repeat for $\phi_1 = .6$, $\phi_2 = -.3$, and $\theta_1 = -.9$.

16.5 Show that the autocorrelation function for the general ARMA(p, q) process is given by

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p} \qquad k \geq q + 1$$

as in Eq. (16.63).

16.6 Suppose that y_t is first-order homogeneous nonstationary, and that $w_t = \Delta y_t$ can be represented by the ARMA(1, 1) model

$$w_t = .9w_{t-1} + \varepsilon_t - .6\varepsilon_{t-1} + 1$$

If $y_t = 0$ for $t = 0$, what is $E(y_t)$ as a function of time?

16.7 Relate the summation operator to the backward shift operator by showing that

$$\Sigma = (1 - B)^{-1} = 1 + B + B^2 + B^3 + \cdots$$

16.8 Refer to the time series for nonfarm inventory investment in Fig. 15.3, its sample autocorrelation function in Fig. 15.4, and its partial autocorrelation function in Fig. 16.8. Can you suggest one or more ARMA(p, q) processes that might have generated that time series?

Right column:

ESTIMATING AND CHECKING TIME-SERIES MODELS

In this chapter we show how the parameters of an ARIMA model are estimated. As we shall see, if the model contains moving average terms, this involves the application of a nonlinear estimation method similar to that described in Chapter 9. Following this we describe *diagnostic checking*, a procedure used to test whether the model has been correctly specified (i.e., whether p, d, and q have been chosen correctly).

The material on estimation is at a somewhat more advanced mathematical level than that in the previous three chapters, so most of the details (which involve matrix notation) have been put in Appendix 17.1. Still, a basic understanding of estimation can be obtained by reading through the chapter and paying special attention to the examples.

17.1 MODEL ESTIMATION

Suppose a tentative specification of the time-series model has been made, i.e., values of p, d, and q have been chosen for the ARIMA model,[1]

$$\phi(B)\Delta^d y_t = \phi(B)w_t = \theta(B)\varepsilon_t \qquad (17.1)$$

with $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$. Now estimates must be obtained for the p autoregressive parameters ϕ_1, . . . , ϕ_p and the q moving average parameters θ_1, . . . , θ_q. As in the case of the

---

[1] We assume for simplicity here that $\delta = 0$, that is, that w_t is measured as a deviation from its mean value.

regression model, we choose parameter values that will minimize the sum of squared differences between the actual time series $w_t = \Delta^d y_t$ and the fitted time series $\hat{w}_t$.

To put this another way, rewrite Eq. (17.1) in terms of the error term series ε_t:[2]

$$\varepsilon_t = \theta^{-1}(B)\phi(B)w_t \qquad (17.2)$$

The objective in estimation is to find a set of autoregressive parameters $(\phi_1, \ldots, \phi_p)$ and a set of moving average parameters $(\theta_1, \ldots, \theta_q)$ that *minimize the sum of squared errors*

$$S(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) = \sum_t \varepsilon_t^2 \qquad (17.3)$$

We denote the sets of parameters that minimize Eq. (17.3) by $(\hat{\phi}_1, \ldots, \hat{\phi}_p)$ and $(\hat{\theta}_1, \ldots, \hat{\theta}_q)$, and the residuals associated with these parameter values by $\hat{\varepsilon}_t$, so that $\hat{\varepsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)w_t$. Thus,

$$S(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q) = \sum_t \hat{\varepsilon}_t^2 \qquad (17.4)$$

This estimation can be difficult if moving average terms are present, since then Eq. (17.2) is nonlinear in the parameters. For this reason an iterative method of nonlinear estimation must be used in the minimization of Eq. (17.3). In addition, the first error term in the series, ε_1, depends on the past and unobservable values $w_0, w_{-1}, \ldots, w_{-p+1}$ and $\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1}$. Thus some method must be used to initialize the series (i.e., choose numbers for these unobservable values) before the nonlinear estimation process is applied.

After the parameters of the model have been estimated, a procedure of *diagnostic checking* is used to test whether the initial specification of the model was correct. We would expect the *residuals* $\hat{\varepsilon}_t$, $t = 1, \ldots, T$, to resemble closely the true errors ε_t, which, by assumption, are uncorrelated. A diagnostic check is used to test whether these residuals are indeed uncorrelated. If they are not, we would want to respecify the model (i.e., choose new values for p, d, and q), estimate this new model, and perform another diagnostic check to determine if it has been correctly specified. Once the model has been checked to satisfaction, it can be used for forecasting future movements in the time series.

Let us examine the estimation procedure in more detail. We assume that a total of $T + d$ observations are available for the homogeneous nonstationary time series of order d, y_t, and we denote these observations as $y_{-d+1}, \ldots, y_0, y_1, \ldots, y_T$. After differencing this series d times, we obtain a stationary series w_t with T observations $w_1, \ldots, w_T$. The problem is to estimate the parameters for the ARMA(p, q) model which has been specified for the series w_t. To do this, we

[2] It should be more clear now why we were concerned with the invertibility of $\theta(B)$ in Appendix 16.1.

utilize the fact that (by assumption) the error terms $\varepsilon_1, \ldots, \varepsilon_T$ are all *normally distributed* and independent, with mean 0 and variance σ_ε^2. Then the *conditional log likelihood function* associated with the parameter values $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_\varepsilon)$ is given by

$$L = -T \log \sigma_\varepsilon - \frac{S(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)}{2\sigma_\varepsilon^2} \qquad (17.5)$$

We say that L is the *conditional* logarithmic likelihood function because the sum of squared errors S depends on the past and unobservable values $w_0, w_{-1}, \ldots, w_{-p+1}, \varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-q+1}$. This can be seen by writing the equation for the first observable error term ε_1 in the expanded form of the ARMA model:

$$\varepsilon_1 = w_1 - \phi_1 w_0 - \phi_2 w_{-1} - \cdots - \phi_p w_{-p+1} + \theta_1 \varepsilon_0 + \cdots + \theta_q \varepsilon_{-q+1} \qquad (17.6)$$

Setting aside for the moment the problem of determining the past values of w_t and ε_t, Eq. (17.5) makes it clear that the maximum-likelihood estimate of the model's parameters is given by the *minimization* of the sum of squared residuals S. Thus, under the assumption of normally distributed errors, the maximum-likelihood estimate is the same as the *least-squares* estimate.

17.1.1 Initialization of the Series

Because the sum-of-squares function $S(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and thus the likelihood function L are both conditional on the past unobservable values of w_t and ε_t ($w_0, \ldots, w_{-p+1}$ and $\varepsilon_0, \ldots, \varepsilon_{-q+1}$), the least-squares estimates that we obtain depend on the choice of values made for $w_0, w_{-1}, \ldots$, etc. For this reason we must choose initial starting values for $w_0, w_{-1}, \ldots$, to be used in the minimization of the conditional sum-of-squares function.

The most common solution to this problem, and the one we recommend, is to set $w_0, \ldots, w_{-p+1}$ and $\varepsilon_0, \ldots, \varepsilon_{-q+1}$ equal to their *unconditional expected values*. The unconditional expected values of $\varepsilon_0, \ldots, \varepsilon_{-q+1}$ are all 0, and, if $\delta = 0$, the unconditional expected values of $w_0, \ldots, w_{-p+1}$ are 0 as well. This solution will provide a reasonably good approximation to the correct procedure if the actual values of $\phi_1, \ldots, \phi_p$ are not very close to 1 and if the number of observations T is large relative to p and q.[3]

17.1.2 Nonlinear Estimation of Model Parameters

Our estimation problem is to find values of the parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ that minimize the sum of squared errors S. Assuming that the initialization

[3] An alternative method of initializing the series is to determine *conditional expected values* for $w_0, \ldots, w_{-p+1}$, that is, values that are conditional on the observed values of $w_1, \ldots, w_T$ and the estimated values of $\varepsilon_1, \ldots, \varepsilon_T$. This procedure is technically difficult and its benefits may not be substantial. We recommend using the unconditional expected values for $w_0, \ldots, w_{-p+1}$, that is, setting them equal to 0 (when $\delta = 0$).

tion of the series is based, as we suggest, on the unconditional expected values (which are all 0) of $w_0, \ldots, w_{-p+1}$ and $\varepsilon_0, \ldots, \varepsilon_{-q+1}$, the time bounds would be $t = 1$ to T. Thus the problem is to pick $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q$ to minimize

$$S = \sum_{t=1}^{T} [\varepsilon_t | \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q] \qquad (17.7)$$

Now suppose that the model was purely autoregressive, i.e., was of the form

$$\phi(B)w_t = \varepsilon_t \qquad (17.8)$$

or

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \varepsilon_t \qquad (17.9)$$

Observe that since Eq. (17.9) is of the general form

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \varepsilon_t \qquad (17.10)$$

it can be estimated simply as a linear regression. Although for a purely autoregressive model the estimation process is essentially a linear regression, the problem is more difficult if the model contains a moving average component as well. In that case we can represent the model as

$$\theta^{-1}(B)\phi(B)w_t = \varepsilon_t \qquad (17.11)$$

Clearly, this "regression equation" is nonlinear in the parameters and cannot be estimated by a simple application of ordinary least squares. However, it can be estimated by a general iterative nonlinear estimation routine. The process is nearly identical to that discussed in Section 9.5 and used in standard nonlinear regression programs.

The nonlinear estimation process uses the first two terms in a Taylor series expansion to linearize Eq. (17.11) around an initial guess for the parameter values. A linear regression is then performed on this linearized equation, least-squares estimates of $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are obtained, and a new linearization of Eq. (17.11) is made around these estimates. Again a linear regression is performed, a second set of parameter estimates is obtained, and a new linearization of Eq. (17.11) is made around this second set of estimates. This process is repeated iteratively until convergence occurs, i.e., until the estimates of the parameters do not change after repeated iterations. In Appendix 17.1 we discuss the actual mechanics of this process in detail.

17.1.3   Obtaining an Initial Guess for the Parameter Values

Before a nonlinear estimation can be performed on Eq. (17.11), an initial guess must be made for the parameter values. Convergence of the estimation process

may be faster if the initial guess is a good one, i.e., close to the "true" parameter values. On the other hand, if the initial guess is very poor, it is possible that the iterative process may not converge at all.

The sample autocorrelation function can sometimes be used to help produce the initial guess. As one might expect, this may work for a low-order time-series model, but it can be virtually useless if the model is at all complicated. For example, if the series w_t is modeled as first-order autoregressive, one need only look at the sample value of ρ_1. If that is, say, .9, a reasonable first guess for ϕ_1 is $\phi_{1,0} = .9$. If, however, our model for w_t is complex, this inspection method is unlikely to produce much useful information.

Even if we cannot determine the initial guess by simply inspecting a correlogram, we can still use the numerical values for the sample autocorrelation function to obtain the initial guess. As we demonstrated in the last chapter, the theoretical autocorrelation function can be related to the theoretical parameter values through a series of equations. If these equations are inverted, they can be used to solve for the parameter values *in terms of the autocorrelation function*. This is straightforward in the case of a purely autoregressive model. As an example, consider the autoregressive process of order p, and recall from Eq. (16.47) that the difference equation for its autocorrelation function is given by

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$$

Using the fact that $\rho_k = \rho_{-k}$, we can rewrite this equation as a set of p simultaneous linear equations relating the parameters $\phi_1, \ldots, \phi_p$ to $\rho_1, \ldots, \rho_p$:

$$
\begin{aligned}
\rho_1 &= \phi_1 \quad\;\; + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 \quad\;\; + \cdots + \phi_p \rho_{p-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p
\end{aligned}
\qquad (17.12)
$$

Using these *Yule-Walker equations* to solve for the parameters $\phi_1, \ldots, \phi_p$ in terms of the *estimated* values of the autocorrelation function, we arrive at the *Yule-Walker estimates* of the parameters. These estimates can be used to provide a reasonable first guess for the parameter values.[4] This first guess is, however, of limited value, since the purely autoregressive model can be estimated by ordinary least squares.

If the time-series model contains a moving average part, the Yule-Walker equations that relate the values of the autocorrelation function to the values of the parameters will not be linear. Recall, for example, that the process MA(1)

---

[4] Writing Eq. (17.12) in matrix notation

$$\rho = P\phi$$

we can solve for ϕ as simply $\phi = P^{-1}\rho$.

has the autocorrelation function

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

Suppose in this example that $\rho_1 = .4$ in the sample autocorrelation function. Then

$$\theta_1 = \frac{-1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1} = \frac{-1 \pm .6}{.8} \tag{17.13}$$

Thus the first estimate for θ_1 is -2 or $-.5$. Since invertibility necessitates that $|\theta_1| < 1$, we select the value $\theta_{1,0} = -.5$ for our first guess in the nonlinear estimation process. Unfortunately, the solution for the θ's in terms of the ρ's becomes more difficult as the moving average order q becomes larger. In fact, to get initial estimates for the model MA(q), it is necessary to solve q simultaneous nonlinear equations. As a result, we often try several initial guesses and see whether our estimates converge to the same final parameter values.

One might ask why parameter values based on the Yule-Walker equations are not sufficient for practical purposes. This would eliminate the use of the nonlinear estimation method. One reason is that the sample autocorrelation function is only an *estimate* of the actual autocorrelation function and thus is subject to error. In fact, for small samples the sample autocorrelation function will be biased (downward) from the true autocorrelation function. A second reason is that the sample autocorrelation function does not contain as much information as the actual time series. To use as much information as possible in the estimation of the model's parameters, we calculate our final estimates based on the actual time series.

17.2 DIAGNOSTIC CHECKING

After a time-series model has been specified and its parameters have been estimated, one must test whether the original specification was correct. This process of diagnostic checking usually involves two steps. First, the autocorrelation function for the simulated series (i.e., the time series generated by the model) can be compared with the sample autocorrelation function of the original series. If the two autocorrelation functions seem very different, some doubt may be cast on the validity of the model and a respecification may be in order. If the two autocorrelation functions are not markedly different (and this will most often be the case), one can analyze the *residuals* of the model.

Remember that we have assumed that the random error terms ε_t in the actual process are normally distributed and *independent*. Then if the model has been specified correctly, the *residuals* $\hat{\varepsilon}_t$ should resemble a white noise process. In particular, we would expect the residuals to be *nearly uncorrelated* with each other, so that a *sample autocorrelation function* of the residuals would be close to 0 for displacement $k \geq 1$.

Recall that the residuals of the model are

$$\hat{\varepsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)w_t \tag{17.14}$$

Let us denote the sample autocorrelation function (for displacement k) of the residuals as $\hat{r}_k$. It is calculated by

$$\hat{r}_k = \frac{\sum\limits_t \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum\limits_t \hat{\varepsilon}_t^2} \tag{17.15}$$

As we mentioned in Chapter 15, a very convenient test, based on statistical results obtained by Box and Pierce, can be applied to this sample autocorrelation function.[5] *If the model is correctly specified, then for large displacements k (for example, $k > 5$ for low-order models) the residual autocorrelations $\hat{r}_k$ are themselves uncorrelated, normally distributed random variables with mean 0 and variance $1/T$*, where T is the number of observations in the time series. This fact makes it possible to devise a simple diagnostic test.

Consider the statistic Q composed of the first K residual autocorrelations $\hat{r}_1$, . . . , $\hat{r}_K$:[6]

$$Q = T \sum_{k=1}^{K} \hat{r}_k^2 \tag{17.16}$$

This statistic is a sum of squared independent normal random variables, each with mean 0 and variance $1/T$ and is therefore itself approximately distributed as *chi-square* (see Chapter 2). We say "approximately" because the first few autocorrelations r_1, r_2, etc., will have a variance slightly less than $1/T$ and may themselves be correlated. Box and Pierce demonstrate that the approximation is quite close and that the statistic Q will be distributed as $\chi^2(K - p - q)$, i.e., chi square with $K - p - q$ degrees of freedom.[7] Therefore, a statistical hypothesis test of model accuracy can be performed by comparing the observed value of Q with the appropriate points from a chi-square table.

[5] G. E. P. Box and D. A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, vol. 65, December 1970.

[6] For low-order models, K equal to 15 or 20 is sufficient.

[7] In Chapter 15 we said that the Q statistic is chi square with K degrees of freedom. Note, however, that that was in reference to a test of the hypothesis that the *original data series* (as opposed to the residuals from our estimated ARMA model) is white noise. For the original data series, $p = q = 0$.

Suppose, for example, that we have specified an ARMA(1, 1) model for a series w_t, that the model has been estimated, and that the statistic Q is calculated to be 31.5 with $K = 20$. From a chi-square table we see that the 90 percent point for $K - p - q = 18$ degrees of freedom is 26.0, and the 95 percent point is 28.9. Thus the statistic Q is too large and we can reject the model, since the probability that the residuals are not white noise is at least 95 percent. Suppose that a new model, ARMA(2, 2), is specified and estimated, and the statistic Q is now 22.0, again with $K = 20$. From the chi-square table we see that the 90 percent point for 16 degrees of freedom is 23.5. Thus we need not reject the hypothesis that the residuals are white, and this second model would be acceptable.[8] To determine the "best" specification, we might want to specify and estimate some other ARMA models to see whether a lower chi-square statistic can be obtained.

If the calculated value of Q is between the 90 and 95 percent points of the chi-square tail, some doubt would be thrown on the model. At the very least a second test should be applied. This second test would involve observing the individual values of $\hat{r}_k$ for all k between, say, $K/4$ and K (in our example, between $k = 5$ and $k = 20$). Since these $\hat{r}_k$ are normal with variance $1/T$, we can test to see if they are all within two or three standard deviations from their means of 0. If several of the $\hat{r}_k$ are larger than $2/\sqrt{T}$ (two standard deviations of the normal variable), evidence exists that the model is misspecified. In addition the evidence might suggest how the model should be respecified. For example, if for an ARMA(2, 1) model $\hat{r}_3$ is very much larger than $2/\sqrt{T}$, this would indicate that the model should be respecified with the inclusion of a third-order moving average term.

In constructing a time-series model one often estimates several alternative specifications. It may be the case that two or more specifications pass the diagnostic checks described above. In this case additional tests must be used to determine the "best" specification. One test is to compare the "simulated series" (i.e., the time series generated by the model) for each specification with the original series. The specification that yields the smallest rms simulation error would then be retained. However, unless one specification has a markedly lower rms error, we suggest retaining all the specifications (that pass the diagnostic checks) and choosing among them based on their *forecasting* performance. The generation and evaluation of forecasts from a time-series model are discussed in the next chapter.

Example 17.1 Interest Rates   In the last two chapters we began analyzing a time series of monthly data for the interest rate on 3-month U.S. government Treasury bills from the beginning of 1950 through June 1988, and a

[8] Note that this chi-square test is a "weak" hypothesis test. A value of Q below the 90 percent point on the chi-square distribution indicates that it is not necessary to reject the hypothesis that the residuals are white, since the probability that the hypothesis is true is less than 90 percent. It is thus only an *indirect* test of the hypothesis that the residuals are not white.

time series of monthly data for U.S. hog production from the beginning of 1960 through the end of 1967. Let us now estimate some alternative ARIMA models for these two time series. We will begin with the interest rate series. In order to allow for long enough lags, we will estimate models using a sample period of February 1951 to June 1988.

Review the sample autocorrelation functions in Figs. 15.8, 15.10, and 15.12 for the series undifferenced, differenced once, and differenced twice. We explained that these autocorrelation functions suggest that the series is first-order homogeneous nonstationary, i.e., can be modeled as ARIMA(p, 1, q). But as we discussed in Chapter 16, a specification for p and q is difficult to determine from the sample autocorrelations. A low-order model, such as ARIMA(2, 1, 2) may suffice, but the fact that the sample autocorrelations remain significantly different from zero even for large lags suggests that it may be necessary to estimate models of much higher order.

We begin with a low-order specification and estimate an ARIMA(2, 1, 2) model. The result is

ARIMA(2, 1, 2):

$$(1 - .7834B + .0944B^2) \Delta y_t = .0115 + (1 - .3899B - .3170B^2)\varepsilon_t$$
(17.17)

$$R^2 = .161 \qquad \chi^2(4, 36) = 123.32$$

Note that while the R^2 of this equation is low, this does not necessarily mean that the specification is a poor one. Remember that the R^2 measures fit in terms of the dependent variable of the regression, which in this case is the *monthly change* in the interest rate. A more revealing statistic is the chi square, which is equal to 123.32. With 32 degrees of freedom (36 lags minus 4 estimated AR and MA parameters), this value is far above the critical 95 percent level. Thus one can conclude (with at least 95 percent certainty) that the residuals from the ARIMA(2, 1, 2) model are autocorrelated, and higher-order terms are needed for the model.

As a next step, we increase the number of parameters somewhat and estimate an ARIMA(4, 1, 4) model. The result is

ARIMA(4, 1, 4):

$$(1 - .2542B + .0993B^2 - .3487B^3 - .3758B^4) \Delta y_t$$
$$= .0106 + (1 + .1620B - .0930B^2 - .3949B^3 - .6387B^4)\varepsilon_t \quad (17.18)$$

$$R^2 = .204 \qquad \chi^2(8, 36) = 85.12$$

While the chi-square statistic has dropped, it is still highly significant, leading us to reject this specification.

We now try specifications that are of much higher order, beginning with ARIMA(12, 1, 9):

ARIMA(12, 1, 9):

$$(1 + .0747B + .3600B^2 + .0821B^3 + .1673B^4 - .4822B^5 + .2760B^6$$
$$- .0272B^7 - .1201B^8 - .3429B^9 + .1431B^{10} - .2679B^{11}$$
$$+ .0856B^{12}) \Delta y_t = .0106 + (1 + .5072B + .3478B^2 + .1414B^3$$
$$+ .0786B^4 - .3933B^5 - .1676B^6 - .1118B^7 - .1891B^8 - .2835B^9)\varepsilon_t$$
$$(17.19)$$

$$R^2 = .296 \quad \chi^2(21, 36) = 26.70$$

The chi-square statistic has dropped considerably, to 26.70. However, there are now only 15 degrees of freedom (36 lags minus 21 estimated parameters), and this value is significant at the 95 percent level (though barely so). We must now consider the possibility that we have included too many parameters; a somewhat sparser specification that saves degrees of freedom might still account for almost as much of the variation in the change in interest rates.

As a next step, we reduce the number of moving average terms, and estimate an ARIMA(12, 1, 3) model and an ARIMA(12, 1, 2) model. The results are shown below:

ARIMA(12, 1, 3):

$$(1 + .2812B + .1335B^2 - .2649B^3 + .2526B^4 - .1463B^5 + .2321B^6$$
$$+ .1394B^7 + .0675B^8 - .2427B^9 + .0049B^{10} - .0561B^{11}$$
$$+ .1649B^{12}) \Delta y_t = .0109 + (1 + .7143B + .2084B^2 - .3142B^3)\varepsilon_t$$
$$(17.20)$$

$$R^2 = .289 \quad \chi^2(15, 36) = 32.03$$

ARIMA(12, 1, 2):

$$(1 + .4211B + .4811B^2 - .0928B^3 + .2139B^4 - .0777B^5 + .2512B^6$$
$$+ .1490B^7 + .1340B^8 - .1556B^9 - .0272B^{10} - .1171B^{11}$$
$$+ .1559B^{12}) \Delta y_t = .0109 + (1 + .8562B + .6257B^2)\varepsilon_t$$
$$(17.21)$$

$$R^2 = .294 \quad \chi^2(14, 36) = 28.16$$

Of these, the ARIMA(12, 1, 2) model looks most promising. The chi-square statistic is 28.16, which, with $36 - 14 = 22$ degrees of freedom, is insignificant even at the 90 percent level. In practice, sets of *ex post* forecasts should be produced using several alternative models so that their perfor-

mance can be compared. We will produce forecasts using the (12, 1, 2) model in the next chapter.

Example 17.2 Hog Production   Let us now turn to the series for monthly hog production. Recall from the last chapter that we suggested that an appropriate model for the series might be

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)(1 - B^{12})y_t = \varepsilon_t$$

This model was estimated over the period January 1960 to December 1967, with the results

$$(1 + .6681B + .2015B^2 - .1298B^3)(1 - B)(1 - B^{12})y_t = .0014 + \varepsilon_t$$
$$(17.22)$$

$$R^2 = .365 \quad \chi^2(3, 20) = 12.83$$

The model is acceptable and will be used to forecast in the next chapter. The reader might wonder, however, whether another model specification might provide a better fit to the data. Perhaps, for example, the addition of moving average terms would improve the model. To test this, we estimated a model that includes first- and second-order moving average terms:

$$(1 + .6626B + .3945B^2 - .0179B^3)(1 - B)(1 - B^{12})y_t$$
$$= .0015 + (1 + .0168B - .2191B^2)\varepsilon_t \quad (17.23)$$

$$R^2 = .349 \quad \chi^2(5, 20) = 13.01$$

Inclusion of the moving average terms results in a slightly lower value for the R^2. Of greater importance, however, is the fact that the estimated values of ϕ_1, ϕ_2, and ϕ_3 add up to a number greater than 1. The result is a nonstationary model for a series that we believe to be stationary. We would thus reject this model, and retain the model of Eq. (17.22).

APPENDIX 17.1 Nonlinear Estimation of ARIMA Models

In this appendix we examine the iterative nonlinear estimation process in more detail. Let the vector $\boldsymbol{\beta}$ represent the $p + q$ parameters $(\boldsymbol{\theta}, \boldsymbol{\phi})$ that we wish to

estimate. We wish to choose numerical values for $\boldsymbol{\beta}$ which minimize

$$S(\boldsymbol{\beta}) = \sum_{t=1}^{T} [\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}]^2 = \sum_{t=1}^{T} [\varepsilon_t]^2 \qquad (A17.1)$$

The notation $[\varepsilon_t]$ means that the errors are conditional on the values of $\mathbf{w}$ and $\boldsymbol{\beta}$. Now expand $[\varepsilon_t]$ in a Taylor series around some initial guess $\boldsymbol{\beta}_0$ for the parameter values:

$$[\varepsilon_t] = [\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}_0] + \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) \frac{\partial [\varepsilon_t]}{\partial \beta_i} \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

$$+ \frac{1}{2} \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0})^2 \frac{\partial^2 [\varepsilon_t]}{\partial \beta_i^2} \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \cdots + \qquad (A17.2)$$

Here $\beta_{i,0}$ is the value of the first guess for the parameter β_i, and thus is the ith component of the vector $\boldsymbol{\beta}_0$. We will approximate $[\varepsilon_t]$ by the first two terms of this Taylor series expansion. When we let

$$x_{i,t} = -\frac{\partial [\varepsilon_t]}{\partial \beta_i} \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \qquad (A17.3)$$

and

$$[\varepsilon_{t,0}] = [\varepsilon_t | \mathbf{w}, \boldsymbol{\beta}_0] \qquad (A17.4)$$

it follows that (approximately)

$$[\varepsilon_t] = [\varepsilon_{t,0}] - \sum_{i=1}^{p+q} (\beta_i - \beta_{i,0}) x_{i,t} \qquad (A17.5)$$

which can be rewritten as

$$[\varepsilon_{t,0}] + \sum_{i=1}^{p+q} \beta_{i,0} x_{i,t} = \sum_{i=1}^{p+q} \beta_i x_{i,t} + [\varepsilon_t] \qquad (A17.6)$$

The left-hand side of Eq. (A17.6) can be thought of as a composite dependent variable, which will have different numerical values for $t = 1, 2, \ldots, T$. (Note that $[\varepsilon_{t,0}]$ is just the value of the error term at time t, given the first guess $\boldsymbol{\beta}_0$.) On the right-hand side of Eq. (A17.6) are $p + q$ independent variables (multiplied by the $p + q$ unknown parameters β_i) as well as an additive error term. It should be clear, then, that the parameters β_i can be estimated from Eq. (A17.6) via a linear regression. This ordinary least-squares regression is performed to produce a new estimate of $\boldsymbol{\beta}$, which we call $\hat{\boldsymbol{\beta}}_1$.

Next, using a new Taylor series expansion of $[\varepsilon_t]$ around this $\hat{\boldsymbol{\beta}}_1$, we obtain a new version of Eq. (A17.6) which can also be estimated by ordinary least squares to yield a new estimate $\hat{\boldsymbol{\beta}}_2$. This process is repeated over and over again until

$$\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_{k-1} \approx 0 \qquad (A17.7)$$

We call k the *convergence number*, i.e., the number of iterations required until convergence occurs. $\hat{\boldsymbol{\beta}}_k$ will then be our final estimate of the parameters ϕ_1, $\ldots$, ϕ_p and $\theta_1, \ldots, \theta_q$.

The standard errors and t statistics for our parameter estimates would be calculated from the *last* linearization, as is typically done in a nonlinear estimation procedure. An R^2 can be calculated in the same way, and has similar relevance.

There can be no guarantee that the estimation process described above will converge *at all* to a final estimate of the parameters. It is quite possible for the process to *diverge*, i.e., for the successive estimates $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}, \ldots, \hat{\boldsymbol{\beta}}_k$ to be farther and farther apart. Furthermore, it is also possible for *multiple solutions* to exist; in this case convergence occurs, but different initial guesses result in different final estimates for one or more of the parameters. Whether or not divergence or multiple solutions occur would depend both on the specification of the ARIMA model and on the data to which that specification is fitted.

Should divergence occur, the model can be reestimated using different initial guesses. A new initial guess may result in convergence, although this need not always be the case. If it turns out to be impossible to reach a convergent solution, a new model specification must be chosen.

Even if convergence occurs on the first try, it is wise to test for multiple solutions by reestimating the model with a different initial guess. If multiple solutions do occur, the final estimate should be that which gives the smallest value of the sum of squared errors. This estimate would correspond to the global minimum of the sum of squared errors function, as opposed to one or more local minima.

Example A17.1 Estimation of ARMA(1, 1) As an example of the iterative nonlinear estimation method described above, let us see how the estimation of ϕ_1 and θ_1 in an ARMA(1, 1) model for a stationary series w_t would be carried out. Writing the ARMA(1, 1) model in terms of Eq. (17.11), we have

$$\varepsilon_t = \frac{1 - \phi_1 B}{1 - \theta_1 B} w_t \qquad (A17.8)$$

This equation is nonlinear in θ_1, but it can be approximated by a Taylor series

expansion. This involves calculating the first derivatives of ε_t with respect to ϕ_1 and θ_1, and evaluating them at the initial guess $\phi_{1,0}$ and $\theta_{1,0}$.

$$x_{1,t} = -\left.\frac{\partial \varepsilon_t}{\partial \phi_1}\right|_{\phi_{1,0},\,\theta_{1,0}} = \frac{B}{1 - \theta_{1,0}B}\, w_t \qquad (A17.9)$$

$$x_{2,t} = -\left.\frac{\partial \varepsilon_t}{\partial \theta_1}\right|_{\phi_{1,0},\,\theta_{1,0}} = -\frac{B - \phi_{1,0}B^2}{(1 - \theta_{1,0}B)^2}\, w_t \qquad (A17.10)$$

Numerical time series can be computed for $x_{1,t}$ and $x_{2,t}$ (they will be used to perform the linear regression) over the time period $t = 1$ to T by expanding Eqs. (A17.9) and (A17.10). Taking Eq. (A17.9) for $x_{1,t}$, for example, we get

$$x_{1,t} = \theta_{1,0}x_{1,t-1} + w_{t-1} \qquad (A17.11)$$

Setting $w_0 = x_{1,0} = 0$, we can solve Eq. (A17.11) repeatedly to generate a series for $x_{1,t}$. The first value $x_{1,1}$ would be equal to 0, the second value $x_{1,2}$ would equal w_1, the third value $x_{1,3}$ would equal $\theta_{1,0}w_1 + w_2$, etc. The same thing is done for $x_{2,t}$ by expanding Eq. (A17.10):

$$x_{2,t} = 2\theta_{1,0}x_{2,t-1} - \theta_{1,0}^2 x_{2,t-2} - w_{t-1} + \phi_{1,0}w_{t-2} \qquad (A17.12)$$

We would begin the computation of $x_{2,t}$ by setting

$$x_{2,1} = x_{2,0} = w_0 = w_{-1} = 0 \qquad (A17.13)$$

Thus, the first value $x_{2,1}$ would just be equal to 0, while $x_{2,2}$ would be equal to $-w_1$, $x_{2,3}$ would be equal to $2\theta_{1,0}x_{2,2} - w_2 + \phi_{1,0}w_1$, etc.

Finally, a time series must also be computed for $\varepsilon_{t,0}$ for t ranging from 1 to T. This is done simply by writing

$$\varepsilon_{t,0} = \frac{1 - \phi_{1,0}B}{1 - \theta_{1,0}B}\, w_t \qquad (A17.14)$$

and using the series for w_t to obtain the series for $\varepsilon_{t,0}$. Note that this is just a series of residuals based on the first-guess estimates $\phi_{1,0}$ and $\theta_{1,0}$.

We are now ready to perform a linear regression. Equation (A17.6) becomes

$$\varepsilon_{t,0} + \phi_{1,0}x_{1,t} + \theta_{1,0}x_{2,t} = \phi_1 x_{1,t} + \theta_1 x_{2,t} + \varepsilon_t \qquad (A17.15)$$

Equation (A17.15) is a linear regression equation, and ordinary least squares can be used to estimate ϕ_1 and θ_1. If these estimates of ϕ_1 and θ_1 are significantly different from the first guesses $\phi_{1,0}$ and $\theta_{1,0}$, they are used as *new* first

guesses, and the *entire process is repeated* all over again, to yield *new* estimates for ϕ_1 and θ_1. Again, if the change is significant, then the process is repeated. With luck, the estimates for ϕ_1 and θ_1 will converge after a few iterations, but we have no guarantee of this, nor, in fact, do we even have a guarantee that the process will converge at all.

EXERCISES

17.1 Following the example in Appendix 17.1, show how the estimation of ϕ_1, ϕ_2, and θ_1 for an ARMA(2, 1) model would be carried out. Go through the steps of the Taylor series expansion, show how the data series are generated, and indicate how the linear regressions are performed.

17.2 Suppose that an ARMA(0, 2) model has been estimated for a time series that has been generated by an ARMA(1, 2) process.

(a) How would the diagnostic test indicate that the model has been misspecified?

(b) What will the residual autocorrelations $\hat{r}_k$ look like? What characteristics of these autocorrelations might indicate that ARMA(1, 2) is a more correct specification?

17.3 Repeat Exercise 17.2 for an ARMA(0, 2) model estimated for a time series that has been generated by an ARMA(2, 3) process.

17.4 Suppose that a particular homogeneous nonstationary time series y_t can be modeled as a stochastic process that is ARIMA(1, 1, 1).

(a) How would you calculate the sample autocorrelation functions for y_t and its differences and use them to verify that ARIMA(1, 1, 1) is indeed a proper specification for y_t?

(b) Suppose you did not have access to a computer package for nonlinear estimation. How would you use a *linear regression* to obtain *approximate* estimates of the parameters in the model? (Explain the steps involved clearly.)

17.5 Using data for the 3-month Treasury bill rate (Table 17.1 on pages 514–515) (or some other short-term interest rate), specify and estimate alternative models to those in Example 17.1. Experiment with higher-order ARIMA(p, 1, q) models, and also with ARIMA(p, 2, q) models. How sensitive are your estimates to the choice of sample period?

TABLE 17.1
THREE-MONTH TREASURY BILL RATE

Obs.	3-month rate					
1950.01	1.07	1.12	1.12	1.15	1.16	1.15
1950.07	1.16	1.20	1.30	1.31	1.36	1.34
1951.01	1.34	1.36	1.40	1.47	1.55	1.45
1951.07	1.56	1.62	1.63	1.54	1.56	1.73
1952.01	1.57	1.54	1.59	1.57	1.67	1.70
1952.07	1.81	1.83	1.71	1.74	1.85	2.09
1953.01	1.96	1.97	2.01	2.19	2.16	2.11
1953.07	2.04	2.04	1.79	1.38	1.44	1.60
1954.01	1.18	.97	1.03	.96	.76	.64
1954.07	.72	.92	1.01	.98	.93	1.14
1955.01	1.23	1.17	1.28	1.59	1.45	1.41
1955.07	1.60	1.90	2.07	2.23	2.25	2.54
1956.01	2.41	2.32	2.25	2.60	2.61	2.49
1956.07	2.31	2.60	2.84	2.90	2.99	3.21
1957.01	3.11	3.11	3.08	3.06	3.06	3.29
1957.07	3.16	3.37	3.53	3.58	3.29	3.04
1958.01	2.44	1.54	1.30	1.13	.91	.83
1958.07	.91	1.69	2.44	2.63	2.67	2.77
1959.01	2.82	2.70	2.80	2.95	2.84	3.21
1959.07	3.20	3.38	4.04	4.05	4.15	4.49
1960.01	4.35	3.96	3.31	3.23	3.29	2.46
1960.07	2.30	2.30	2.48	2.30	2.37	2.25
1961.01	2.24	2.42	2.39	2.29	2.29	2.33
1961.07	2.24	2.39	2.28	2.30	2.48	2.60
1962.01	2.72	2.73	2.72	2.73	2.68	2.73
1962.07	2.92	2.82	2.78	2.74	2.83	2.87
1963.01	2.91	2.92	2.89	2.90	2.92	2.99
1963.07	3.18	3.32	3.38	3.45	3.52	3.52
1964.01	3.52	3.53	3.54	3.47	3.48	3.48
1964.07	3.46	3.50	3.53	3.57	3.64	3.84
1965.01	3.81	3.93	3.93	3.93	3.89	3.80
1965.07	3.84	3.84	3.92	4.03	4.09	4.38
1966.01	4.59	4.65	4.59	4.62	4.64	4.50
1966.07	4.80	4.96	5.37	5.35	5.32	4.96
1967.01	4.72	4.56	4.26	3.84	3.60	3.54
1967.07	4.21	4.27	4.42	4.56	4.73	4.97
1968.01	5.00	4.98	5.17	5.38	5.66	5.52
1968.07	5.31	5.09	5.19	5.35	5.45	5.96
1969.01	6.14	6.12	6.02	6.11	6.04	6.44
1969.07	7.00	6.98	7.09	7.00	7.24	7.82
1970.01	7.87	7.13	6.63	6.51	6.84	6.68
1970.07	6.45	6.41	6.13	5.91	5.28	4.87
1971.01	4.44	3.70	3.38	3.86	4.14	4.75
1971.07	5.40	4.94	4.69	4.46	4.22	4.01
1972.01	3.38	3.20	3.73	3.71	3.69	3.91
1972.07	3.98	4.02	4.66	4.74	4.78	5.07
1973.01	5.41	5.60	6.09	6.26	6.36	7.19
1973.07	8.01	8.67	8.29	7.22	7.83	7.45

TABLE 17.1
THREE-MONTH TREASURY BILL RATE (Continued)

Obs.	3-month rate					
1974.01	7.77	7.12	7.96	8.33	8.23	7.90
1974.07	7.55	8.96	8.06	7.46	7.47	7.15
1975.01	6.26	5.50	5.49	5.61	5.23	5.34
1975.07	6.13	6.44	6.42	5.96	5.48	5.44
1976.01	4.87	4.88	5.00	4.86	5.20	5.41
1976.07	5.23	5.14	5.08	4.92	4.75	4.35
1977.01	4.62	4.67	4.60	4.54	4.96	5.02
1977.07	5.19	5.49	5.81	6.16	6.10	6.07
1978.01	6.44	6.45	6.29	6.29	6.41	6.73
1978.07	7.01	7.08	7.85	7.99	8.64	9.08
1979.01	9.35	9.32	9.48	9.46	9.61	9.06
1979.07	9.24	9.52	10.26	11.70	11.79	12.04
1980.01	12.00	12.86	15.20	13.20	8.58	7.07
1980.07	8.06	9.13	10.27	11.62	13.73	15.49
1981.01	15.02	14.79	13.36	13.69	16.30	14.73
1981.07	14.95	15.51	14.70	13.54	10.86	10.85
1982.01	12.28	13.48	12.68	12.70	12.09	12.47
1982.07	11.35	8.68	7.92	7.71	8.07	7.94
1983.01	7.86	8.11	8.35	8.21	8.19	8.79
1983.07	9.08	9.34	9.00	8.64	8.76	9.00
1984.01	8.90	9.09	9.52	9.69	9.83	9.87
1984.07	10.12	10.47	10.37	9.74	8.61	8.06
1985.01	7.76	8.27	8.52	7.95	7.48	6.95
1985.07	7.08	7.14	7.10	7.16	7.24	7.10
1986.01	7.07	7.06	6.56	6.06	6.15	6.21
1986.07	5.83	5.53	5.21	5.18	5.35	5.53
1987.01	5.43	5.59	5.59	5.64	5.66	5.67
1987.07	5.69	6.04	6.40	6.13	5.69	5.77
1988.01	5.81	5.66	5.70	5.91	6.26	6.46

Source: Citibase, Series FYGM3.

FORECASTING WITH TIME-SERIES MODELS

Once a time-series model has been estimated and checked, it can be used for forecasting. In this chapter we explain how to use the general ARIMA model

$$\phi(B)\Delta^d y_t = \theta(B)\varepsilon_t \qquad (18.1)$$

to obtain a forecast of y_t for period $T + l$ (that is, l periods ahead, with $l \geq 1$). We denote this forecast by $\hat{y}_T(l)$, and call it the *origin-T forecast for lead time l*. We assume for now that the true parameters of the model are known and examine the properties both of the forecast and of the forecast error. Later we will see how imperfect knowledge of the true parameter values increases the forecast error.

We begin this chapter by discussing the basis for making forecasts, after which we go through the steps of actually computing a forecast. Then we discuss the nature of forecast errors, showing how forecast confidence intervals can be computed. In order to give the reader an understanding of the characteristics of time-series forecasts, we examine in detail the properties of the forecasts of some simple ARIMA models. Finally, we present two examples in which we generate forecasts for an interest rate and for hog production using the time-series models estimated at the end of the last chapter.

18.1 MINIMUM MEAN SQUARE ERROR FORECAST

Our objective in forecasting is to predict future values of a time series subject to as little error as possible. For this reason we consider the optimum forecast to be

that forecast which has the *minimum mean square forecast error*. Since the forecast error is a random variable, we minimize the *expected value*. Thus we wish to choose our forecast $\hat{y}_T(l)$ so that $E[e_T^2(l)] = E\{[y_{T+l} - \hat{y}_T(l)]^2\}$ is minimized. We show that this forecast is given by the *conditional expectation* of y_{T+l}, that is, by

$$\hat{y}_T(l) = E(y_{T+l}|y_T, y_{T-1}, \ldots, y_1) \qquad (18.2)$$

To prove that the minimum mean square error forecast is given by Eq. (18.2), we begin by rewriting the ARIMA model in Eq. (18.1) above as

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t \qquad (18.3)$$

since $\Delta = 1 - B$, as explained in Chapter 16. Therefore,

$$y_t = \phi^{-1}(B)(1 - B)^{-d}\theta(B)\varepsilon_t = \psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad (18.4)$$

Here we have expressed the ARIMA model as a purely moving average process of infinite order.[1] Then

$$y_{T+l} = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_l \varepsilon_T + \psi_{l+1} \varepsilon_{T-1} + \cdots$$

$$= \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1}\varepsilon_{T+1} + \sum_{j=0}^{\infty} \psi_{l+j}\varepsilon_{T-j} \qquad (18.5)$$

In Eq. (18.5) we have divided the infinite sum into two parts, the second part beginning with the term $\psi_l \varepsilon_T$ and thus describing information up to and including time period T.

Of course the forecast $\hat{y}_T(l)$ can be based only on information available up to time T. Our objective is to compare this forecast with the actual value y_{T+l} as expressed in Eq. (18.5). To do so, we write the forecast as a weighted sum of those error terms *which we can estimate*, namely, $\varepsilon_T, \varepsilon_{T-1}, \ldots$. Then, the desired forecast is

$$\hat{y}_T(l) = \sum_{j=0}^{\infty} \psi_{l+j}^* \varepsilon_{T-j} \qquad (18.6)$$

where the weights ψ_{l+j}^* are to be chosen optimally so as *to minimize the mean*

[1] Any ARIMA process can be equivalently expressed as purely moving average or as purely autoregressive. We could have, for example, rewritten Eq. (18.3) as $\phi(B)(1 - B)^d\theta^{-1}(B)y_t = \varepsilon_t$, or $\xi(B)y_t = \varepsilon_t$. This is a purely autoregressive process of infinite order. The reason that we do not originally specify the ARIMA process as purely autoregressive or purely moving average (of infinite order) is that we would then have an infinite number of parameters to estimate.

square forecast error. We can now write an expression for the forecast error, $e_T(l)$, using Eqs. (18.5) and (18.6):

$$e_T(l) = y_{T+l} - \hat{y}_T(l) = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1}\varepsilon_{T+1}$$

$$+ \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)\varepsilon_{T-j} \qquad (18.7)$$

Since by assumption $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$, the mean square forecast error is

$$E[e_T^2(l)] = (\psi_0^2 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_\varepsilon^2 + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_\varepsilon^2 \qquad (18.8)$$

Clearly this expression is minimized by setting the "optimum" weights ψ_{l+j}^* equal to the true weights ψ_{l+j}, for $j = 0, 1, \ldots$. But then our optimum forecast $\hat{y}_T(l)$ is just the conditional expectation of y_{T+l}. This can be seen by taking the conditional expectation of y_{T+l} in Eq. (18.5). The expected values of $\varepsilon_{T+l}, \ldots, \varepsilon_{T+1}$ are all 0, while the expected values of ε_T, $\varepsilon_{T-1}, \ldots$, are just the residuals from the estimated equation. Thus we have

$$\hat{y}_T(l) = \sum_{j=0}^{\infty} \psi_{l+j}\hat{\varepsilon}_{T-j} = E(y_{T+l}|y_T, \ldots, y_1) \qquad (18.9)$$

This provides the basic principle for calculating forecasts from our ARIMA models. Now we apply this principle to the actual computation of forecasts.

18.2 COMPUTING A FORECAST

The computation of the forecast $\hat{y}_T(l)$ can be done recursively using the estimated ARIMA model. This involves first computing a forecast one period ahead, using this forecast to compute a forecast two periods ahead, and continuing until the l-period forecast has been reached. Let us write the ARIMA(p, d, q) model as

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} + \delta \qquad (18.10)$$

with

$$y_t = \Sigma^d w_t \qquad (18.11)$$

To compute the forecast $\hat{y}_T(l)$, we begin by computing the *one-period* forecast of w_t, $\hat{w}_T(1)$. To do so, we write Eq. (18.10) with the time period modified:

$$w_{T+1} = \phi_1 w_T + \cdots + \phi_p w_{T-p+1} + \varepsilon_{T+1} - \theta_1 \varepsilon_T - \cdots - \theta_q \varepsilon_{T-q+1} + \delta$$

$$(18.12)$$

We then calculate our forecast $\hat{w}_T(1)$ by taking the conditional expected value of w_{T+1} in Eq. (18.12):

$$\hat{w}_T(1) = E(w_{T+1}|w_T, \ldots) = \phi_1 w_T + \cdots + \phi_p w_{T-p+1}$$

$$- \theta_1 \hat{\varepsilon}_T - \cdots - \theta_q \hat{\varepsilon}_{T-q+1} + \delta \qquad (18.13)$$

where the $\hat{\varepsilon}_T$, $\hat{\varepsilon}_{T-1}$, etc., are observed residuals. Note that the expected value of ε_{T+1} is 0. Now using the one-period forecast $\hat{w}_T(1)$, we can obtain the *two-period* forecast $\hat{w}_T(2)$:

$$\hat{w}_T(2) = E(w_{T+2}|w_T, \ldots)$$

$$= \phi_1 \hat{w}_T(1) + \phi_2 w_T + \cdots + \phi_p w_{T-p+2} - \theta_2 \hat{\varepsilon}_T - \cdots - \theta_q \hat{\varepsilon}_{T-q+2} + \delta$$

$$(18.14)$$

The two-period forecast is then used to produce the three-period forecast, and so on, until the l-period forecast $\hat{w}_T(l)$ is reached:

$$\hat{w}_T(l) = \phi_1 \hat{w}_T(l - 1) + \cdots + \phi_l w_T + \cdots + \phi_p w_{T-p+l}$$

$$- \theta_l \hat{\varepsilon}_T - \cdots - \theta_q \hat{\varepsilon}_{T-q+l} + \delta \qquad (18.15)$$

Note that if $l > p$ and $l > q$, then this forecast will be

$$\hat{w}_T(l) = \phi_1 \hat{w}_T(l - 1) + \cdots + \phi_p \hat{w}_T(l - p) \qquad (18.16)$$

Once the differenced series w_t has been forecasted, a forecast can be obtained for the original series y_t simply by applying the summation operation to w_t, that is, by summing w_t d times. Suppose, for example, that $d = 1$. Then our l-period forecast of y_t would be given by

$$\hat{y}_T(l) = y_T + \hat{w}_T(1) + \hat{w}_T(2) + \cdots + \hat{w}_T(l) \qquad (18.17)$$

On the other hand, if the model for y_t were ARIMA with $d = 2$, then the l-period forecast $\hat{y}_T(l)$ would be given by

$$\hat{y}_T(l) = y_T + [\Delta y_T + \hat{w}_T(1)] + [\Delta y_T + \hat{w}_T(1) + \hat{w}_T(2)] + \cdots$$

$$+ [\Delta y_T + \hat{w}_T(1) + \cdots + \hat{w}_T(l)]$$

$$= y_T + l \Delta y_T + l\hat{w}_T(1) + (l - 1)\hat{w}_T(2) + \cdots + \hat{w}_T(l) \qquad (18.18)$$

Here the summation operator has been applied twice. The procedure is similar for larger values of d.

18.3 THE FORECAST ERROR

As we saw before, if we express the ARIMA model as a purely moving average process of infinite order, the forecast error l periods ahead is given by

$$e_T(l) = y_{T+l} - \hat{y}_T(l) = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1} \varepsilon_{T+1} \quad (18.19)$$

Remember that the weights ψ_j are determined from

$$\psi(B) = \phi^{-1}(B)(1 - B)^{-d}\theta(B) \quad (18.20)$$

We assume that the model parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are known exactly and therefore the weights $\psi_0, \psi_1, \ldots,$ are also known exactly. In this case the *variance* of the forecast error is given by

$$E[e_T^2(l)] = (\psi_0^2 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_\varepsilon^2 \quad (18.21)$$

Therefore, the algebraic form for the forecast error variance depends on the particular ARIMA specification that has been adopted. In the next section we examine the forecast error in more detail for some simple ARIMA models. For now, however, there are two things that the reader should observe.

First, we know from the definition of $\psi(B)$ above that $\psi_0 = 1$.[2] Therefore, for *any* ARIMA specification, we know that the forecast error *one period* ahead is just

$$e_T(1) = \varepsilon_{T+1} \quad (18.22)$$

and this has variance σ_ε^2. Thus the forecast error variance one period ahead is the variance of the error term.

Second, we must keep in mind the fact that our calculation of the forecast error was based on the assumption that we knew the parameter values $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ with certainty. But the parameters are estimated via a nonlinear least-squares regression, and the estimates are random variables with means and variances. Therefore the *actual* forecast error variance will be *larger* than the variance calculated above. To determine exactly *how much larger*, we must know the variances of the parameter estimates in the ARIMA model. Because the parameters are estimated nonlinearly, however, the best we could do, in fact, would be to calculate standard errors based on the *last iteration* of the nonlinear estimation procedure.

The difficulty here is that the standard errors for the linearization in the last (or any particular) iteration are not "true" estimates of the actual standard errors for the parameter values. As a practical matter, one has the choice of using these standard errors in the calculation of the forecast error variance or ignoring them and simply calculating the forecast error variance based on Eq. (18.21) above.

---

[2] Remember that the ARIMA model (with 0 mean) is $w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$. The only unlagged term on the right-hand side is ε_t (which has a weight of 1). Thus ψ_0 must equal 1 in Eq. (18.4).

18.4 FORECAST CONFIDENCE INTERVALS

Before we can calculate a confidence interval for our forecast, we need an estimate $\hat{\sigma}_\varepsilon^2$ for the variance of the disturbance term. This estimate would logically be based on the sum of squared residuals $S(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$ obtained after final estimates of the parameters have been obtained:

$$\hat{\sigma}_\varepsilon^2 = \frac{S}{T - p - q} = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}{T - p - q} \quad (18.23)$$

Here $T - p - q$ is the number of degrees of freedom in the linear regression. We see from Eq. (18.21) and the fact that $\psi_0 = 1$ that a *confidence interval of n standard deviations* around a forecast l periods ahead would be given by

$$C_n = \hat{y}_T(l) \pm n \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} \hat{\sigma}_\varepsilon \quad (18.24)$$

As expected, this interval gets larger as the lead time l becomes larger, although the exact pattern depends on the weights ψ_j.

Forecasts of y_t, together with a typical 66 percent confidence interval ($n = 1$) and 95 percent confidence interval ($n = 2$), are shown for a hypothetical ARMA model ($d = 0$) in Fig. 18.1. Note that the forecasts (denoted by crosses) first are increasing but then decline to the constant mean level of the series. We know that the forecast will approach the mean of the series as the lead time l becomes

FIGURE 18.1
Forecasts and confidence intervals for a stationary ARMA process.
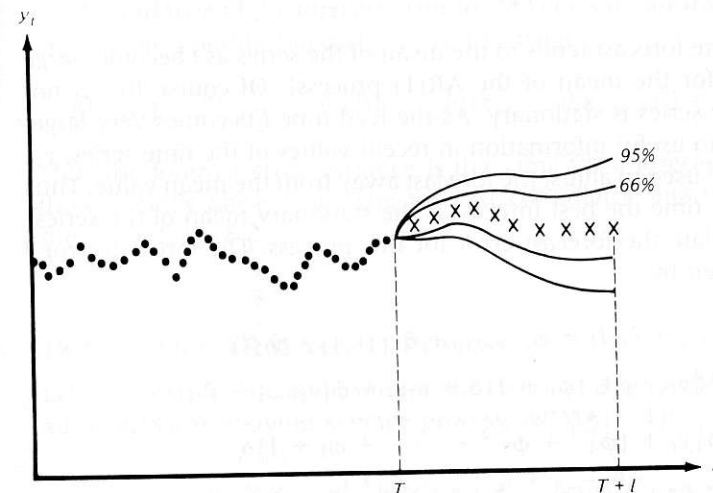
large because the process is stationary. The confidence intervals, of course, increase as the forecast lead time becomes longer.

18.5 PROPERTIES OF ARIMA FORECASTS

We now examine the properties of the forecasts derived from some simple ARIMA models. In all the cases that follow we assume that the parameters of the particular ARIMA model are known with certainty.

18.5.1 The AR(1) Process

Let us begin with the stationary first-order autoregressive process, AR(1):

$$y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t \tag{18.25}$$

For this process the one-period forecast is

$$\hat{y}_T(1) = E(y_{T+1}|y_T, \dots, y_1) = \phi_1 y_T + \delta \tag{18.26}$$

Similarly,

$$\hat{y}_T(2) = \phi_1 \hat{y}_T(1) + \delta = \phi_1^2 y_T + (\phi_1 + 1)\delta \tag{18.27}$$

And the l-period forecast is

$$\hat{y}_T(l) = \phi_1^l y_T + (\phi_1^{l-1} + \phi_1^{l-2} + \cdots + \phi_1 + 1)\delta \tag{18.28}$$

Note that in the limit as l becomes large, the forecast converges to the value

$$\lim_{l \to \infty} \hat{y}_T(l) = \delta \sum_{j=0}^{\infty} \phi_1^j = \frac{\delta}{1 - \phi_1} = \mu_y \tag{18.29}$$

We see, then, that the forecast tends to the mean of the series as l becomes large [recall Eq. (16.23) for the mean of the AR(1) process]. Of course this is not surprising, since the series is stationary. As the lead time l becomes very large, there is essentially no useful information in recent values of the time series, y_T, y_{T-1}, etc., that can be used to adjust the forecast away from the mean value. Thus for a very large lead time the best forecast is the stationary mean of the series.

Let us now calculate the forecast error for this process. The forecast error l periods ahead is given by

$$\begin{aligned} e_T(l) = y_{T+l} - \hat{y}_T(l) &= \phi_1 y_{T+l-1} + \delta + \varepsilon_{T+l} - \hat{y}_T(l) \\ &= \phi_1^2 y_{T+l-2} + (\phi_1 + 1)\delta + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} - \hat{y}_T(l) \\ &\cdots \cdots \cdots \\ &= \phi_1^l y_T + (\phi_1^{l-1} + \phi_1^{l-2} + \cdots + \phi_1 + 1)\delta \\ &\quad + \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \cdots + \phi_1^{l-1}\varepsilon_{T+1} - \hat{y}_T(l) \end{aligned}$$

Now substituting Eq. (18.28) for $\hat{y}_T(l)$, we get

$$e_T(l) = \varepsilon_{T+l} + \phi_1 \varepsilon_{T+l-1} + \cdots + \phi_1^{l-1}\varepsilon_{T+1} \tag{18.30}$$

which has a variance

$$E[e_T^2(l)] = (1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2l-2})\sigma_\varepsilon^2 \tag{18.31}$$

Note that this forecast error variance increases (nonlinearly) as l becomes larger.

18.5.2 The MA(1) Process

Now let us examine the simple first-order moving average process, MA(1):

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{18.32}$$

The one-period forecast for this process is

$$\hat{y}_T(1) = E(y_{T+1}|y_T, \dots, y_1) = \delta - \theta_1 \hat{\varepsilon}_T \tag{18.33}$$

where $\hat{\varepsilon}_T$ is the actual residual from the current (and most recent) observation. On the other hand, the l-period forecast, for $l > 1$, is just

$$\hat{y}_T(l) = E(y_{T+l}|y_T, \dots, y_1) = E(\delta + \varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1}) = \delta \tag{18.34}$$

This is also as expected, since the process MA(1) has a memory of only one period. Thus recent data are of no help in making a forecast two or more periods ahead, and the best forecast is the mean of the series, δ.

The variance of the forecast error for MA(1) is σ_ε^2 for the one-period forecast, and for the l-period forecast, $l > 1$, it is given by

$$E[e_T^2(l)] = E\{[y_{T+l} - \hat{y}_T(l)]^2\} = E[(\varepsilon_{T+l} - \theta_1 \varepsilon_{T+l-1})^2] = (1 + \theta_1^2)\sigma_\varepsilon^2 \tag{18.35}$$

Thus the forecast error variance is the same for a forecast two periods ahead, three periods ahead, etc. The forecast confidence intervals would appear as shown in Fig. 18.2.

18.5.3 The ARMA(1, 1) Process

Let us now calculate and examine the forecasts generated by the simplest mixed autoregressive–moving average process, ARMA(1, 1):

$$y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{18.36}$$
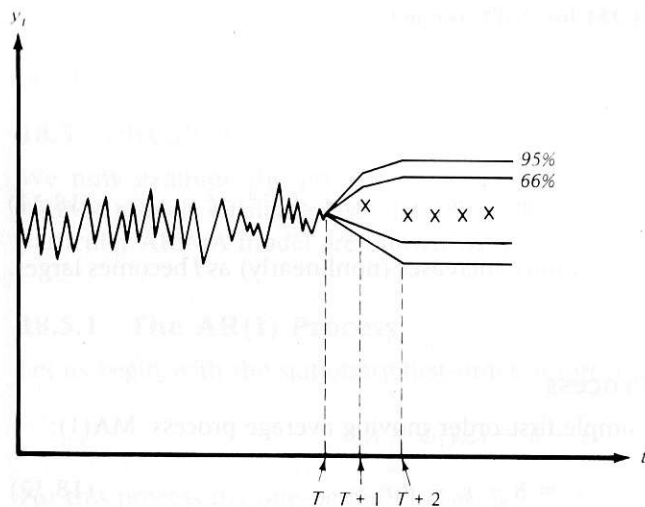
FIGURE 18.2
Forecasts and confidence intervals for an MA(1) process.

The one-period forecast for the ARMA(1, 1) model is given by

$$\hat{y}_T(1) = E(\phi_1 y_T + \delta + \varepsilon_{T+1} - \theta_1 \varepsilon_T) = \phi_1 y_T + \delta - \theta_1 \hat{\varepsilon}_T \qquad (18.37)$$

The two-period forecast is

$$\hat{y}_T(2) = E(\phi_1 y_{T+1} + \delta + \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1}) = \phi_1 \hat{y}_T(1) + \delta$$
$$= \phi_1^2 y_T + (\phi_1 + 1)\delta - \phi_1 \theta_1 \hat{\varepsilon}_T \qquad (18.38)$$

Finally, the l-period forecast is

$$\hat{y}_T(l) = \phi_1 \hat{y}_T(l - 1) + \delta$$
$$= \phi_1^l y_T + (\phi_1^{l-1} + \cdots + \phi_1 + 1)\delta - \phi_1^{l-1}\theta_1 \hat{\varepsilon}_T \qquad (18.39)$$

Note that the limiting value of the forecast as l becomes large is again the mean of the series:

$$\lim_{l \to \infty} \hat{y}_T(l) = \frac{\delta}{1 - \phi_1} = \mu_y \qquad (18.40)$$

Examining these forecasts for different lead times, we see that the current disturbance helps to determine the one-period forecast and, in turn, serves as a starting point from which the remainder of the forecast profile, which is autoregressive in character, decays toward the mean $\delta/(1 - \phi_1)$.

The fact that forecasts from ARMA models approach the (constant) mean value of the series as the lead time becomes large indicates a limitation of these models. As we will see in the examples in this chapter and the next, time-series models are best for short-term forecasting. For a long forecasting horizon, a structural econometric model is likely to be more useful.

18.5.4 The ARI(1, 1, 0) Process

Now we examine a simple nonstationary process, the integrated autoregressive process ARI(1, 1, 0):

$$w_t = \phi_1 w_{t-1} + \delta + \varepsilon_t \qquad (18.41)$$

with

$$w_t = \Delta y_t = y_t - y_{t-1} \qquad (18.42)$$

Forecasts for y_t are related to forecasts of the differenced series w_t as follows:

$$\hat{y}_T(1) = y_T + \hat{w}_T(1) \qquad (18.43)$$

and

$$\hat{y}_T(l) = y_T + \hat{w}_T(1) + \cdots + \hat{w}_T(l) \qquad (18.44)$$

Since the differenced process w_t is AR(1), its forecasts are given by

$$\hat{w}_T(l) = \phi_1^l w_T + (\phi_1^{l-1} + \phi_1^{l-2} + \cdots + \phi_1 + 1)\delta$$
$$= \phi_1^l y_T - \phi_1^l y_{T-1} + (\phi_1^{l-1} + \cdots + \phi_1 + 1)\delta \qquad (18.45)$$

Then the one-period forecast for y_t is

$$\hat{y}_T(1) = y_T + \phi_1(y_T - y_{T-1}) + \delta = (1 + \phi_1)y_T - \phi_1 y_{T-1} + \delta \qquad (18.46)$$

The two-period forecast for y_t is

$$\hat{y}_T(2) = y_T + \hat{w}_T(1) + \hat{w}_T(2) = \hat{y}_T(1) + \hat{w}_T(2)$$
$$= \hat{y}_T(1) + \phi_1^2 w_T + (\phi_1 + 1)\delta$$
$$= (1 + \phi_1 + \phi_1^2)y_T - (\phi_1 + \phi_1^2)y_{T-1} + (\phi_1 + 1)\delta + \delta \qquad (18.47)$$

A more instructive way to look at this forecast, however, is in terms of its *changes*. Since

$$\hat{w}_T(2) = \phi_1 \hat{w}_T(1) + \delta \qquad (18.48)$$

we can write the forecast $\hat{y}_T(2)$ as

$$\hat{y}_T(2) = \hat{y}_T(1) + \phi_1 \hat{w}_T(1) + \delta \qquad (18.49)$$

Similarly,
$$\hat{y}_T(l) = \hat{y}_T(l-1) + \phi_1 \hat{w}_T(l-1) + \delta \qquad (18.50)$$

Now let us examine the properties of this forecast. Since w_t is an AR(1) process, we know from Eq. (18.29) that

$$\lim_{l \to \infty} \hat{w}_T(l) = \frac{\delta}{1 - \phi_1} \qquad (18.51)$$

Thus as the forecast horizon l becomes large, the forecast profile approaches a straight line with slope $\delta/(1 - \phi_1)$. In other words, as the horizon becomes large, the forecast becomes dominated by the *deterministic drift* of the process. For a short forecast horizon this would not be so. It might have been the case, for example, that the last few differences w_T, w_{T-1}, w_{T-2} were negative although δ was positive, so that the series had an overall upward drift. In this case the short-term forecasts $\hat{w}_T(1)$ and $\hat{w}_T(2)$ might be negative, even though $\hat{w}_T(l)$ would tend toward $\delta/(1 - \phi_1)$ as l became larger. The forecasts for y_t, then, would first be decreasing but then would *change direction*, ultimately approaching a straight line with slope $\delta/(1 - \phi_1)$. This hypothetical ARI(1, 1, 0) forecast is shown graphically in Fig. 18.3.

One thing that becomes immediately clear about ARIMA forecasts is that they are *adaptive*. As can be seen from Fig. 18.3, the forecast makes use of the most recent data and adapts accordingly. Another example of the adaptive nature of

FIGURE 18.3
Hypothetical forecasts for an ARI(1, 1, 0) process.
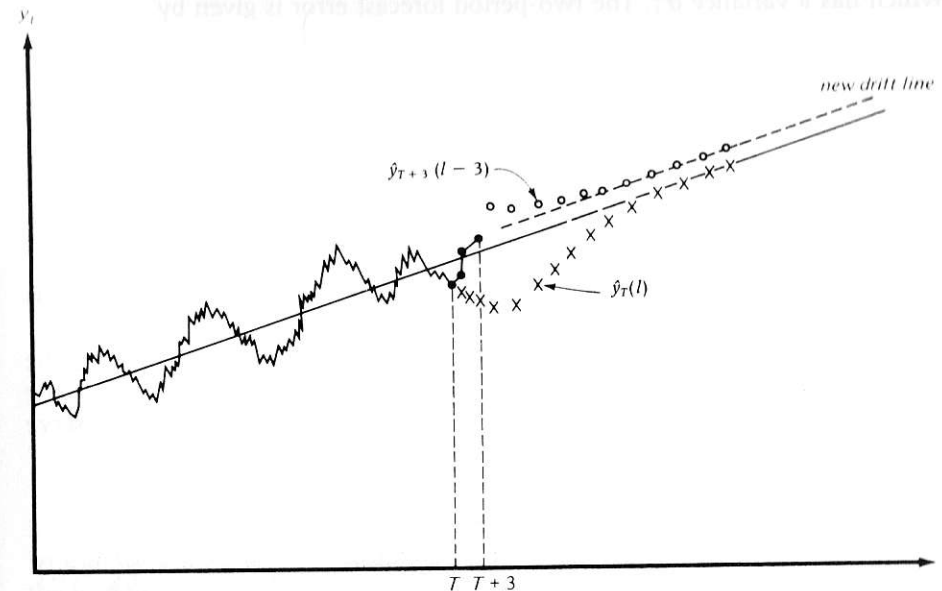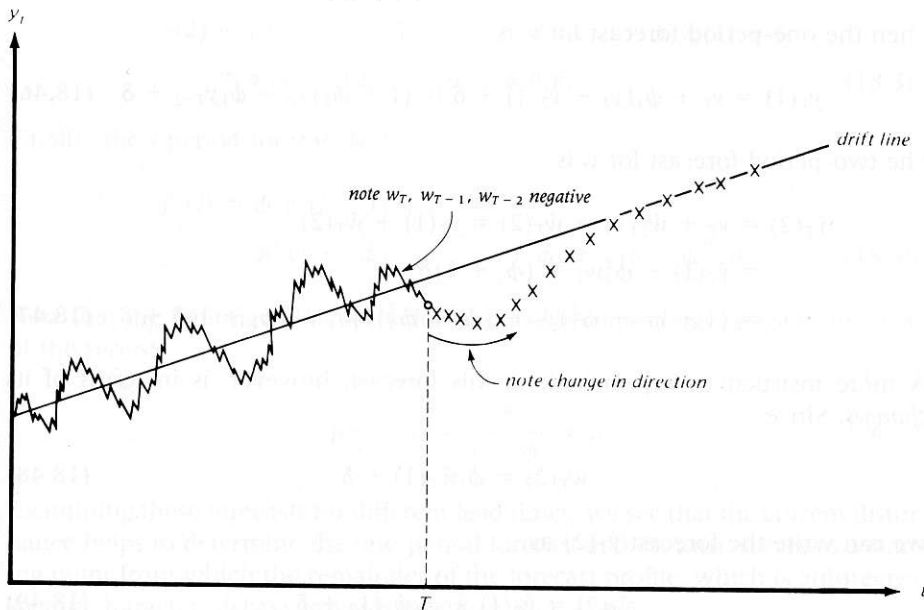
FIGURE 18.4
Adaptive nature of ARI(1, 1, 0) forecast.

ARIMA forecasts is shown in Fig. 18.4. This process is also ARI(1, 1, 0) and is identical to the process in Fig. 18.3 for $t \le T$. The crosses in Fig. 18.4 represent the forecasts made at time T. Now suppose that the series *increases* in periods $T + 1$, $T + 2$, and $T + 3$ and a *new* set of forecasts is made in period $T + 3$. These forecasts are denoted by circles, and, as can be seen in Fig. 18.4, they first *increase* and then *decrease*. Ultimately they will also approach a drift line. This new drift line will have the same slope as before but will be slightly higher as a result of the new data points. What we observe, then, is that the forecast has "adapted" to the new data that became available in periods $T + 1$, $T + 2$, and $T + 3$. Notice that the values of this forecast for a long lead time have adapted as well.

18.5.5 Confidence Intervals for the ARI(1, 1, 0) Forecast

We now calculate the forecast error and its variance for the ARI(1, 1, 0) process, so that we can obtain a forecast confidence interval. As we will see, the forecast confidence interval for y_t is related to the forecast confidence interval for the differenced series w_t.

We begin with the forecast error for the one-period forecast, $\hat{y}_T(1)$:

$$e_T(1) = y_{T+1} - \hat{y}_T(1) = y_T + w_{T+1} - y_T - \hat{w}_T(1)$$
$$= w_{T+1} - \hat{w}_T(1) = \varepsilon_{T+1} \qquad (18.52)$$

which has a variance σ_ε^2. The two-period forecast error is given by

$$e_T(2) = y_{T+2} - \hat{y}_T(2) = y_T + w_{T+1} + w_{T+2} - y_T - \hat{w}_T(1) - \hat{w}_T(2)$$

$$= [w_{T+1} - \hat{w}_T(1)] + [w_{T+2} - \hat{w}_T(2)]$$

$$= (1 + \phi_1)\varepsilon_{T+1} + \varepsilon_{T+2} \tag{18.53}$$

and this has a variance

$$E[e_T^2(2)] = \sigma_\varepsilon^2[(1 + \phi_1)^2 + 1] \tag{18.54}$$

Note that this forecast error (and its variance) is cumulative; i.e., it is equal to the two-period error for $\hat{w}_T(2)$ in *addition* to the one-period error for $\hat{w}_T(1)$. Thus the error in $\hat{y}_T(2)$ is an *accumulation* of the errors in $\hat{w}_T(1)$ and in $\hat{w}_T(2)$. Now observe this cumulative phenomenon in the *l*-period forecast:

$$e_T(l) = [w_{T+1} - \hat{w}_T(1)] + [w_{T+2} - \hat{w}_T(2)] + \cdots + [w_{T+l} - \hat{w}_T(l)]$$

$$= \varepsilon_{T+1} + (\varepsilon_{T+2} + \phi_1\varepsilon_{T+1}) + \cdots + (\varepsilon_{T+l} + \phi_1\varepsilon_{T+l-1} + \cdots + \phi_1^{l-1}\varepsilon_{T+1})$$

$$= (1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{l-1})\varepsilon_{T+1} + (1 + \phi_1 + \cdots + \phi_1^{l-2})\varepsilon_{T+2}$$

$$+ \cdots + (1 + \phi_1)\varepsilon_{T+l-1} + \varepsilon_{T+l}$$

$$= \sum_{i=1}^{l} \varepsilon_{T+i} \sum_{j=0}^{l-i} \phi_1^j \tag{18.55}$$

and this has a variance

$$E[e_T^2(l)] = \sigma_\varepsilon^2 \sum_{i=1}^{l} \left(\sum_{j=0}^{l-i} \phi_1^j\right)^2 \tag{18.56}$$

Thus the error in $\hat{y}_T(l)$ is an accumulation of errors in $\hat{w}_T(1)$, $\hat{w}_T(2)$, . . . , $\hat{w}_T(l)$. This can be seen graphically in Figs. 18.5 and 18.6, which compare confidence intervals for forecasts of the differenced series w_t with confidence intervals for forecasts of y_t. Note the relationship between the forecasts of the differenced series w_t and the forecasts of y_t. w_{T-2} and w_{T-1} are decreasing, and w_T is negative, so that $\hat{w}_T(1)$ and $\hat{w}_T(2)$ are also negative [$\hat{y}_T(1)$ and $\hat{y}_T(2)$ are decreasing], $\hat{w}_T(3)$, $\hat{w}_T(4)$, etc., are positive [$\hat{y}_T(3)$ is larger than $\hat{y}_T(2)$], and finally $\hat{w}_T(l)$ approaches the mean $\delta/(1 - \phi_1)$ as l becomes large [so that $\hat{y}_T(l)$ approaches the drift line]. Observe that the confidence interval for $\hat{y}_T(l)$ grows rapidly, since it must account for the accumulation of forecast errors in the differenced series.

We have examined the forecast properties of only the simplest of ARIMA models, but some of our conclusions apply to more complicated (i.e., higher-order) models. In particular, note that a moving average model of order q has a memory of only q periods, so that the observed data will affect the forecast only
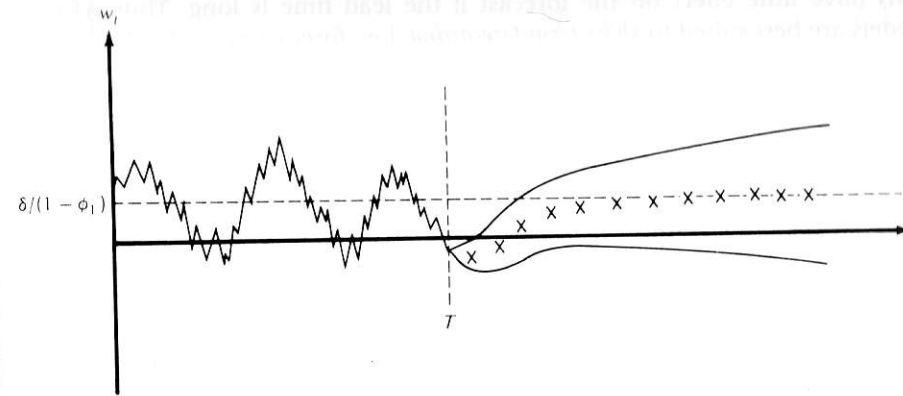
FIGURE 18.5
Confidence interval for $\hat{w}_T(l)$ for ARI(1, 1, 0) process.

if the lead time l is less than q. An autoregressive model has a memory of infinite length, so that all past observations will have some effect on the forecast, even if the lead time l is long. But although all past observations have *some* effect on the forecast, only more recent observations will have a large effect. Thus even with autoregressive (or mixed autoregressive–moving average) models, past observa-

FIGURE 18.6
Confidence interval for $\hat{y}_T(l)$ for ARI(1, 1, 0) process.
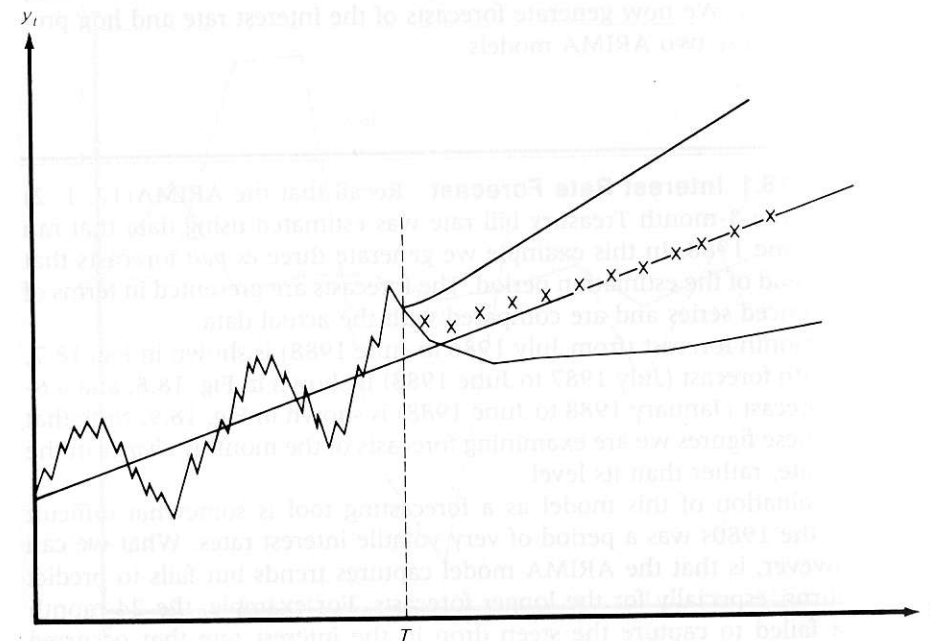
tions have little effect on the forecast if the lead time is long. Thus ARIMA models are best suited to *short-term forecasting*, i.e., forecasting with a lead time l not much longer than $p + q$.

18.6 TWO EXAMPLES

In the last chapter we estimated ARIMA models for two time series. We found that the first series, which consisted of monthly data for the interest rate on 3-month Treasury bills, could be represented using an ARIMA(12, 1, 2) model. The estimated equation is

$$(1 + .4211B + .4811B^2 - .0928B^3 + .2139B^4 - .0777B^5 + .2512B^6$$
$$+ .1490B^7 + .1340B^8 - .1556B^9 - .0272B^{10} - .1171B^{11}$$
$$+ .1559B^{12})\Delta y_t = .0109 + (1 + .8562B + .6257B^2)\varepsilon_t \quad (17.21)$$

The second time series consisted of data on monthly hog production in the United States, which we represented by applying an ARIMA(3, 1, 0) model to a twelfth-differencing of the original series. The estimated version of that model is

$$(1 + .6681B + .2015B^2 - .1298B^3)(1 - B)(1 - B^{12})y_t = .0014 + \varepsilon_t$$

$$(17.22)$$

Recall that the twelfth-differencing $(1 - B^{12})$ accounts for seasonal (annual) cycles in the data. We now generate forecasts of the interest rate and hog production using these two ARIMA models.

Example 18.1 Interest Rate Forecast Recall that the ARIMA(12, 1, 2) model for the 3-month Treasury bill rate was estimated using data that ran through June 1988. In this example we generate three *ex post* forecasts that cover the end of the estimation period. The forecasts are presented in terms of the differenced series and are compared with the actual data.

A 24-month forecast (from July 1986 to June 1988) is shown in Fig. 18.7, a 12-month forecast (July 1987 to June 1988) is shown in Fig. 18.8, and a 6-month forecast (January 1988 to June 1988) is shown in Fig. 18.9. Note that in all of these figures we are examining forecasts of the monthly *change* in the interest rate, rather than its level.

An evaluation of this model as a forecasting tool is somewhat difficult because the 1980s was a period of very volatile interest rates. What we can see, however, is that the ARIMA model captures trends but fails to predict sharp turns, especially for the longer forecasts. For example, the 24-month forecast failed to capture the steep drop in the interest rate that occurred
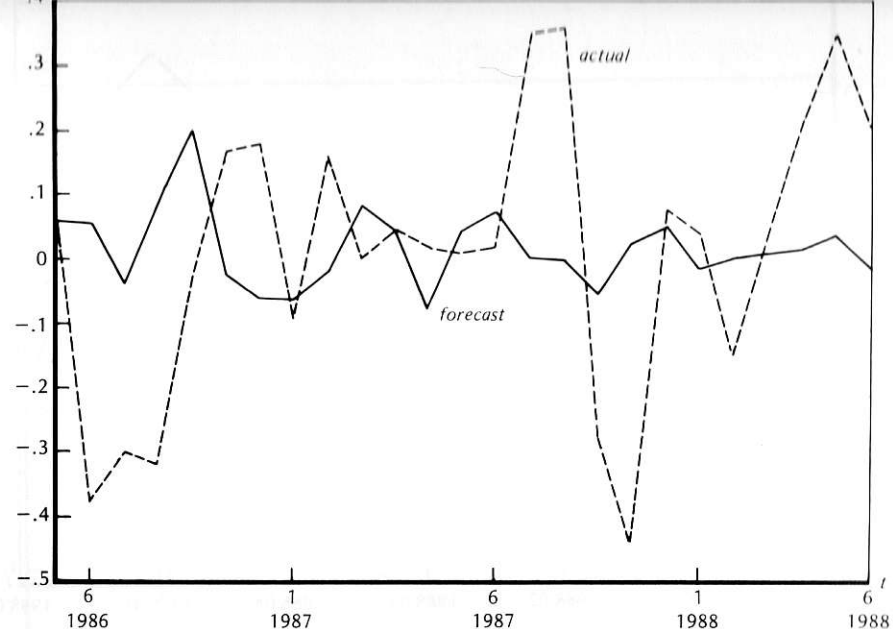
FIGURE 18.7
Monthly changes in Treasury bill rate, 24-month forecast versus actual.

FIGURE 18.8
Monthly changes in Treasury bill rate, 12-month forecast versus actual.

531
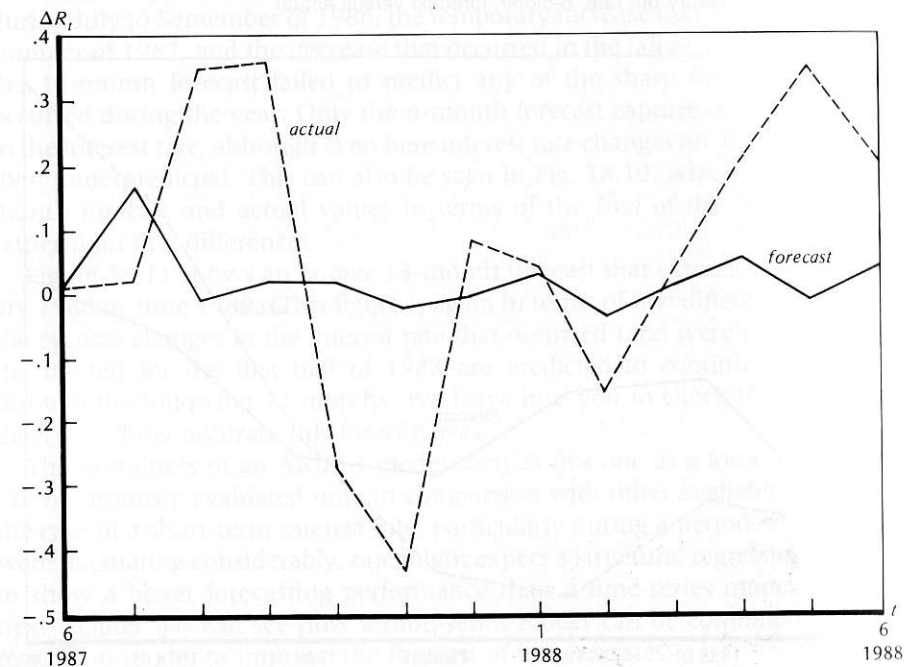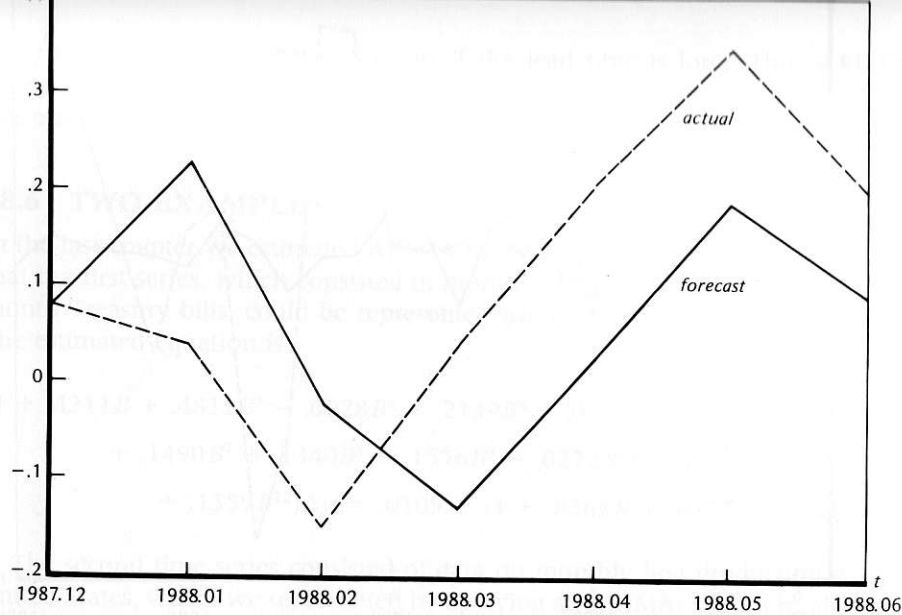
FIGURE 18.9
Monthly changes in Treasury bill rate, 6-month forecast versus actual.

FIGURE 18.10
Three-month Treasury bill rate, 6-month forecast versus actual.
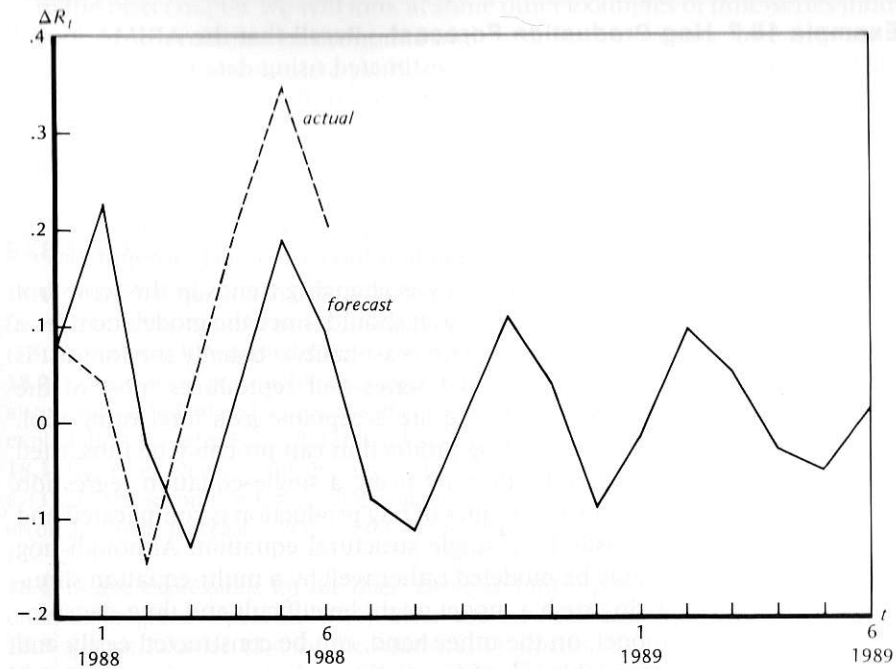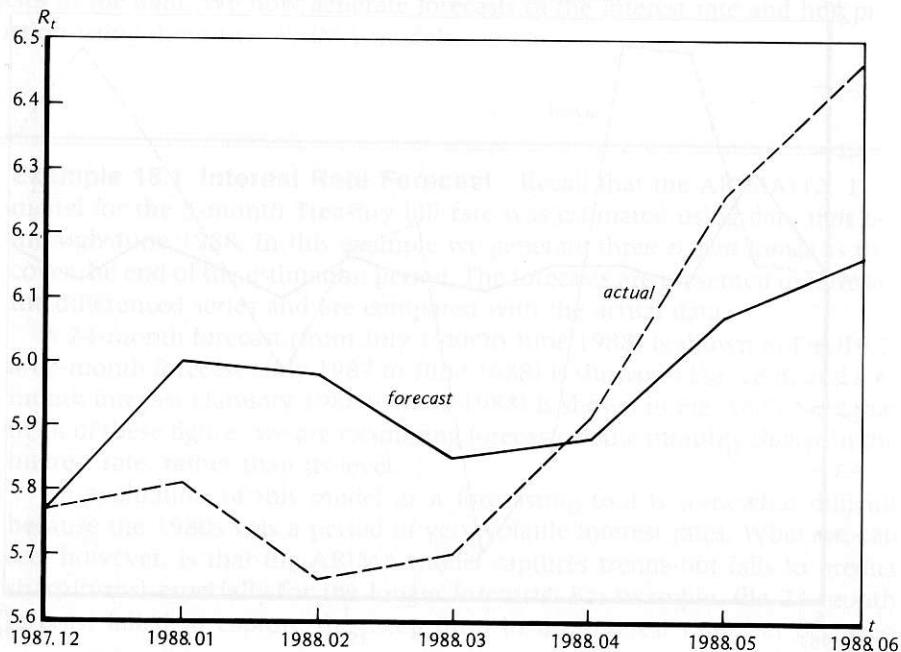
FIGURE 18.11
Monthly changes in Treasury bill rate, *ex ante* forecast.

during July to September of 1986, the temporary increase that occurred in the summer of 1987, and the decrease that occurred in the fall of 1987. Likewise, the 12-month forecast failed to predict any of the sharp movements that occurred during the year. Only the 6-month forecast captures turning points in the interest rate, although even here interest rate changes are first over- and then underpredicted. This can also be seen in Fig. 18.10, which plots the 6-month forecast and actual values in terms of the *level* of the interest rate, rather than first differences.

Figure 18.11 shows an *ex ante* 18-month forecast that extends from January 1988 to June 1989. (This figure is again in terms of first differences.) Here the cyclical changes in the interest rate that occurred (and were predicted by the model) for the first half of 1988 are predicted to continue occurring through the following 12 months. We leave it to you to check the data and determine how accurate this forecast was.

The usefulness of an ARIMA model such as this one as a forecasting tool can be seriously evaluated only in comparison with other available tools. In the case of a short-term interest rate, particularly during a period when rates were fluctuating considerably, one might expect a structural regression model to show a better forecasting performance than a time-series model. In the next chapter we will see how a time-series model can be combined with a regression model to improve the forecast of interest rates.

Example 18.2 Hog Production Forecast   Recall that the ARIMA model for hog production in Eq. (17.22) was estimated using data from the beginning of 1960 to the end of 1967. We generate our forecast out over a 2-year horizon, beginning in January 1968 and ending in January 1970. Since data on hog production are available for this period, we can compare the 25 months of forecasted production with the actual data.

The forecasted and actual series for hog production are shown in Fig. 18.12. Observe that our model has generated forecasts which are quite accurate. The model not only correctly forecasts changing trends in the series but also picks up the broad seasonal cycle (as it should, since the model includes a twelfth-difference of the series to explain seasonality). Usually the forecast is within 10 or 15 percent of the actual series and reproduces most of the turning points. This model would be quite acceptable as a forecasting tool. Unlike our interest rate example, hog production can probably be forecasted better using a time-series model than by using a single-equation regression model. The reason is that the economics of hog production is complicated and cannot be represented easily by a single structural equation. Although hog production could probably be modeled rather well by a multi-equation simulation model, constructing such a model might be difficult and time-consuming. The time-series model, on the other hand, can be constructed easily and quickly and does a reasonable job of forecasting.

FIGURE 18.12
Two-year (25-month) forecast of hog production. Time bounds: January 1968 to January 1970.
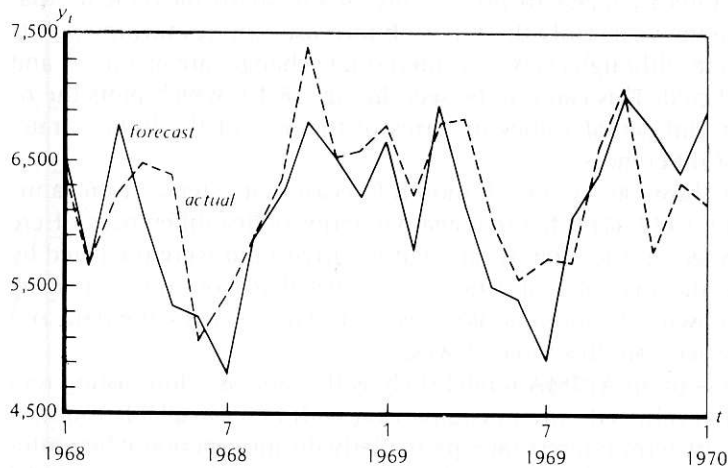
In the next chapter we will look at some other examples of time-series models as applied to problems in economic and business forecasting. In each case we will go through the complete process of specifying, estimating, and checking an ARIMA model, and we will then use the model to produce forecasts. This should provide the reader with more of a feeling for the properties and characteristics of time-series models and forecasts.

EXERCISES

18.1 Write the equation that determines the forecast $\hat{y}_T(l)$ in terms of $\hat{w}_T(1)$, $\hat{w}_T(2)$, . . . , for a third-order homogeneous nonstationary process; i.e., derive the equivalent of Eq. (18.18) for an ARIMA model with $d = 3$.

18.2 Does it seem reasonable that for any ARIMA specification the forecast error variance one period ahead is always the variance of the error term? Offer an intuitive explanation for why Eq. (18.22) must always hold.

18.3 Derive expressions for the one-, two-, and three-period forecasts, $\hat{y}_T(1)$, $\hat{y}_T(2)$, and $\hat{y}_T(3)$, for the second-order moving average process MA(2). What are the variances of the errors for these forecasts? What is the variance of the error for the l-period forecast, with $l > 3$?

18.4 Derive expressions for the one-, two-, and three-period forecasts for the second-order autoregressive process AR(2). What are the error variances of these forecasts?

18.5 Repeat Exercise 18.4 for the ARMA(2, 1) process.

18.6 Suppose that a particular nonstationary time series y_t can be modeled as a stochastic process that is ARIMA(1, 1, 1).

(a) After you have estimated the model's parameters, how would you forecast y_t one period ahead? Express this one-period forecast, $\hat{y}_t(1)$, as a function of observable data. In what sense is this forecast adaptive?

(b) How would you calculate the standard error of the one-period forecast $\hat{y}_t(1)$ *assuming that the parameters of the model are known perfectly*? Note that this is analogous to calculating the standard error of a regression forecast under the assumption that the coefficients β are known perfectly.

(c) What will be the *difference* between the l-period forecast $\hat{y}_t(l)$ and the $(l + 1)$-period forecast $\hat{y}_t(l + 1)$ *when l is very large*?

18.7 In Exercise 17.5 we asked you to estimate alternative ARIMA models for the 3-month Treasury bill rate. Now use your models to generate forecasts comparable to those in Example 18.1. Have you been able to construct a model whose forecasting performance is better?

APPLICATIONS OF TIME-SERIES MODELS

We have seen that econometric model building is in part an art. Even with a simple single-equation model one must make judgments as to which explanatory variables to include, the functional form for the equation, how the statistical fit of the model should be interpreted, and how useful the resulting model is for forecasting or explanation purposes. The situation is much the same with time-series models. It is usually not obvious what the proper specification for an ARIMA model should be. Many different specifications might be reasonable for a single time series and its autocorrelation function, so that sound judgment must be used together with a certain amount of experimentation. As in the regression case, one will often specify and estimate more than one ARIMA model and check each individually. In general, the usefulness of an ARIMA model for forecasting purposes is difficult to ascertain. While confidence intervals can be determined for the model's forecasts, one must still decide whether any significant structural change in the determination of the variable under study might occur and thus alter the future movement of the time series.

In this chapter we present several examples of the construction and use of time-series models. We hope that these examples will help convey a better understanding of the modeling process and will acquaint the reader with the usefulness of time-series models in applied forecasting problems. We will see that time-series models can be used in forecasting applications not only by themselves but also in combination with regression models.

We will begin with a model for an aggregate economic variable, nonfarm inventory investment, and then turn to a model for forecasting seasonal telephone data. One might argue that inventory investment can be better explained by a structural econometric model, but such a model can be difficult and time-consuming to build. The seasonal telephone data that we examine are cyclical, highly fluctuating, and difficult to explain using a structural econometric model, so that a time-series model provides a natural vehicle for forecasting.

As a final application, we show in two examples how it is possible to combine a time-series model with a structural econometric model. To do so, we first construct a regression model and then develop a time-series model for the regression residuals (i.e., for the unexplained noise terms). This combined regression–time-series model is sometimes called a *transfer function* model, and if it is used properly, it can provide a very effective forecasting tool.

19.1 REVIEW OF THE MODELING PROCESS

We begin by briefly reviewing the steps involved in the construction, evaluation, and use of time-series models. One begins with the *specification* of the model. This first requires a decision as to the degree of homogeneity in the time series, i.e., how many times it must be differenced to yield a stationary series. The decision is made by looking at the autocorrelation functions for the series and its differences. (We have seen, however, that the degree of homogeneity is not always obvious.) Then the orders of the moving average and the autoregressive parts of the model must be determined. One can get some guidance from the total and partial sample autocorrelation functions, but often the correct choice will not be clear and several alternative specifications must be estimated.

Once a model (or a group of models) has been specified it must be *estimated*. If the number of observations in the time series is large relative to the order of the model, this estimation process involves a straightforward nonlinear regression. Afterward, one performs a *diagnostic check*. This involves looking at the autocorrelation function of the residuals from the estimated model. A simple chi-square test can be performed to determine whether or not the residuals are themselves uncorrelated. In addition, one should check that the parameter estimates are consistent with *stationarity*, e.g., that the autoregressive parameters sum to a number smaller than 1 in magnitude.

If the model passes the diagnostic check, it must then be *evaluated* to determine its ability to forecast accurately and to provide a better understanding of its forecasting properties. For example, the model may pass a diagnostic check but have a very poor statistical fit, and this would limit its usefulness for forecasting. If the model's estimated parameters have large standard errors, the standard error of forecast will be large.

One means of model evaluation and analysis is to perform a *historical simulation* beginning at different points in time. One can then examine such statistics as the rms simulation error and the Theil inequality coefficient and its decomposition. (See Chapter 12 for a review of these and other model evaluation statistics.) In addition, one can perform an *ex post forecast*, comparing the forecast to actual data to evaluate its performance. This can help the researcher decide how far into the future the model can be used for forecasting. This is extremely important if a time-series model is to be used in conjunction with a structural econometric

model. Typically, the time-series model will provide a better forecast over the very short term, but the structural econometric model will provide a better forecast over the longer term.

19.2 MODELS OF ECONOMIC VARIABLES: INVENTORY INVESTMENT

In this section we construct and examine some time-series models for the level of real (1982 constant dollar) nonfarm inventory investment. This variable is difficult to explain and forecast using structural econometric models, so that the construction of an ARIMA model seems appropriate.

Our sample consists of quarterly data from 1950-1 to 1988-1. In order to allow for sufficient lags in our ARIMA models, estimation will be based on the time period 1952-1 to 1988-1. The time series is shown in Fig. 19.1, and its sample autocorrelation function is shown in Fig. 19.2. When we first examined this time series in Chapter 15, we noted that the sample autocorrelation function exhibits the properties of a stationary series. (After a displacement lag k of 3, it quickly falls toward zero.) In addition, the series itself seems stationary since there are no long-run trends either upward or downward.

Nonetheless, as a check we also difference the series once. The differenced series and its sample autocorrelation function are shown in Figs. 19.3 and 19.4.

FIGURE 19.2
Nonfarm inventory investment: sample autocorrelation function.

FIGURE 19.1
Nonfarm inventory investment (in 1982 constant dollars).

FIGURE 19.3
Inventory investment—first differences (in 1982 constant dollars).
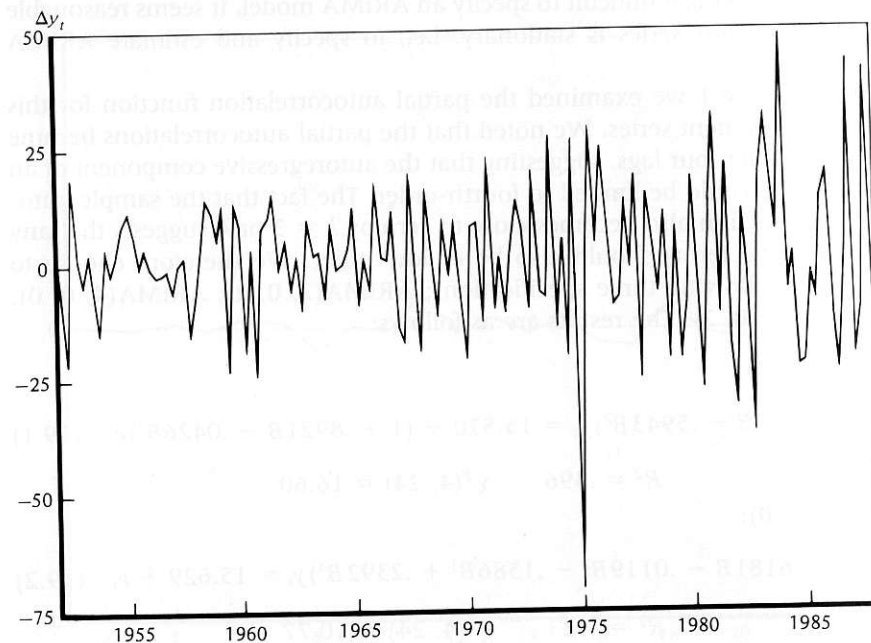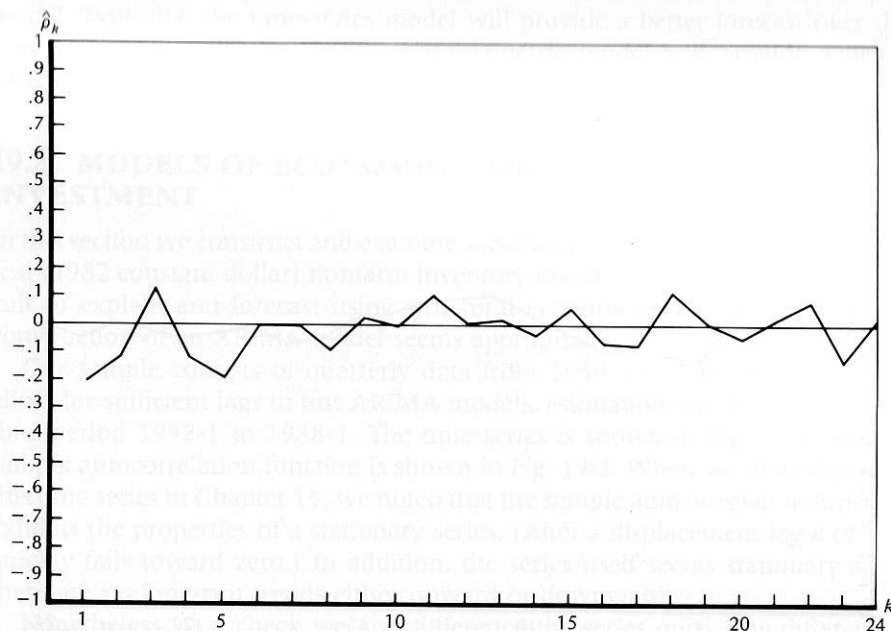
FIGURE 19.4
Inventory investment—first differences: sample autocorrelation function.

Note that the autocorrelation function drops immediately to a value of $-.2$, and then it oscillates between values of roughly $\pm.1$. There is little in the way of a pattern here, making it difficult to specify an ARIMA model. It seems reasonable to assume that our series is stationary, i.e., to specify and estimate ARIMA $(p, 0, q)$ models.

In Example 16.1 we examined the partial autocorrelation function for this inventory investment series. We noted that the partial autocorrelations became close to zero after four lags, suggesting that the autoregressive component of an ARIMA model could be limited to fourth-order. The fact that the sample auto-correlation function also becomes close to zero by $k = 3$ or 4 suggests that any moving average terms should also be of low order. We therefore choose to estimate the following three specifications: ARIMA(2, 0, 2), ARIMA(4, 0, 0), and ARIMA(4, 0, 2). The results are as follows:

ARIMA(2, 0, 2):

$$(1 + .2675B - .5943B^2)y_t = 15.570 + (1 + .8921B - .0426B^2)\varepsilon_t \quad (19.1)$$

$$R^2 = .396 \qquad \chi^2(4, 24) = 16.60$$

ARIMA(4, 0, 0):

$$(1 - .6181B - .0119B^2 - .1586B^3 + .2392B^4)y_t = 15.629 + \varepsilon_t \quad (19.2)$$

$$R^2 = .423 \qquad \chi^2(4, 24) = 10.77$$

ARIMA(4, 0, 2):

$$(1 - .8714B + .5030B^2 - .3814B^3 + .2399B^4)y_t$$

$$= 15.580 + (1 - .2682B + .3792B^2)\varepsilon_t \quad (19.3)$$

$$R^2 = .429 \qquad \chi^2(8, 24) = 7.29$$

All of these chi-square statistics (with 22, 20, and 18 degrees of freedom, respectively) are insignificant even at the 90 percent level, allowing us in each case to accept the hypothesis that the residuals are white noise. It is clear that inventory investment can be described by a low-order ARIMA model. The ARIMA(4, 0, 2) model seems most promising because it has the lowest chi-square statistic, even adjusting for degrees of freedom. Figure 19.5 shows the sample autocorrelation function for the residuals of this model; note that all the autocorrelations are extremely small.

Before proceeding, we estimate an ARIMA(6, 0, 4) model as a means of exploring the possibility that adding more AR and MA terms to the model might improve its fit. The results of that estimation are as follows:

FIGURE 19.5
Autocorrelation function of ARIMA(4, 0, 2) residuals.
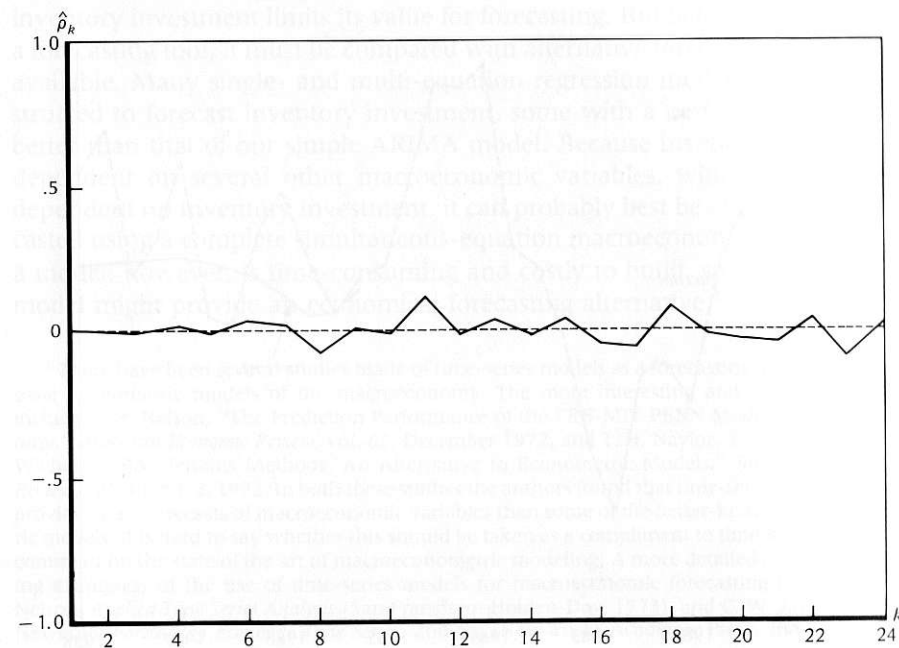
ARIMA(6, 0, 4):

$$(1 - .7495B + .1039B^2 - .2171B^3 + .3833B^4 - .0702B^5 + .0775B^6)y_t$$

$$= 15.637 + (1 - .1463B + .0501B^2 - .0360B^3 + .1571B^4)\varepsilon_t \quad (19.4)$$

$$R^2 = .430 \qquad \chi^2(10, 24) = 7.58$$

There is clearly little benefit to increasing the order of the model. The R^2 is only very slightly higher, and the chi-square statistic has increased; and accounting for degrees of freedom it is much more significant than was the case with the ARIMA(4, 0, 2) model. We will therefore use the ARIMA(4, 0, 2) model to forecast inventory investment.

We first generate a 12-quarter *ex post* forecast from 1985-2 to 1988-1. The forecasted and actual series are shown in Fig. 19.6. Although the forecast follows the overall trend in inventory investment, it does not capture the cyclical fluctuations that occurred during this period. Might a shorter forecast perform better? Figure 19.7 shows a four-quarter *ex post* forecast from 1987-2 to 1988-1. While the forecasted series moves in the same directions as the actual series, the forecast does not capture the extent of the fluctuations. Finally, Fig. 19.8 shows an *ex ante* forecast that extends from 1988-1 to 1990-1. We leave it to you to determine the accuracy of this forecast.

FIGURE 19.6
Twelve-quarter forecast of inventory investment, forecast versus actual.
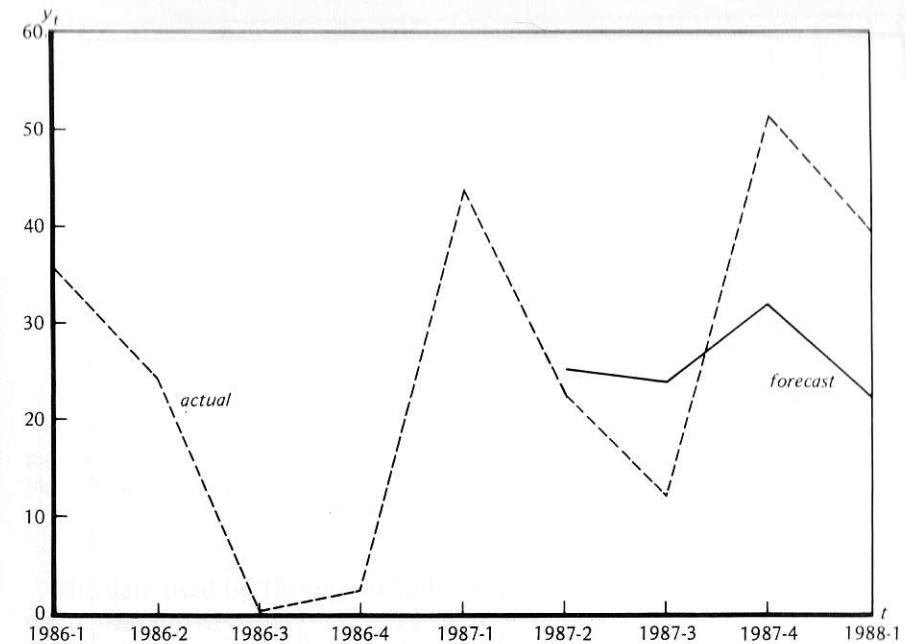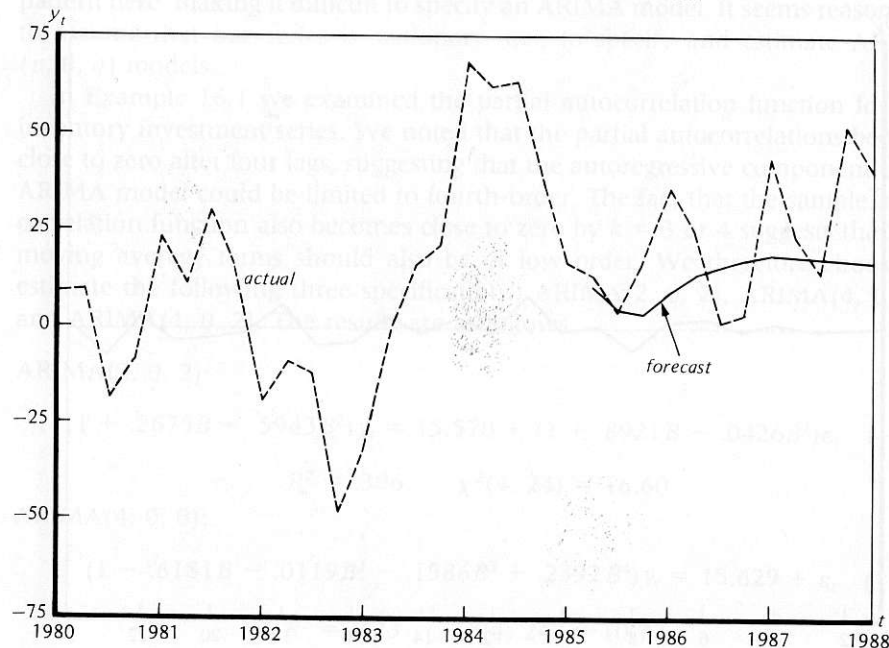
FIGURE 19.7
Four-quarter forecast of inventory investment, forecast versus actual.

The inability of our ARIMA model to predict sharp downturns and upturns in inventory investment limits its value for forecasting. But before it is discarded as a forecasting tool, it must be compared with alternative forecasting tools that are available. Many single- and multi-equation regression models have been constructed to forecast inventory investment, some with a performance not much better than that of our simple ARIMA model. Because inventory investment is dependent on several other macroeconomic variables, which are themselves dependent on inventory investment, it can probably best be explained and forecasted using a complete simultaneous-equation macroeconometric model. Such a model, however, is time-consuming and costly to build, so that a time-series model might provide an economical forecasting alternative.[1]

[1] There have been several studies made of time-series models as a forecasting alternative to large-scale econometric models of the macroeconomy. The more interesting and illuminating studies include C. R. Nelson, "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review*, vol. 62, December 1972, and T. H. Naylor, T. G. Seaks, and D. W. Wichern, "Box-Jenkins Methods: An Alternative to Econometric Models," *International Statistical Review*, vol. 40, no. 2, 1972. In both these studies the authors found that time-series models can often provide better forecasts of macroeconomic variables than some of the better-known large econometric models. It is hard to say whether this should be taken as a compliment to time-series analysis or a comment on the state of the art of macroeconometric modeling! A more detailed discussion (including examples) of the use of time-series models for macroeconomic forecasting is given in C. R. Nelson, *Applied Time Series Analysis* (San Francisco: Holden-Day, 1973), and C. W. J. Granger and P. Newbold, *Forecasting Economic Time Series*, 2nd ed. (New York: Academic Press, 1986).
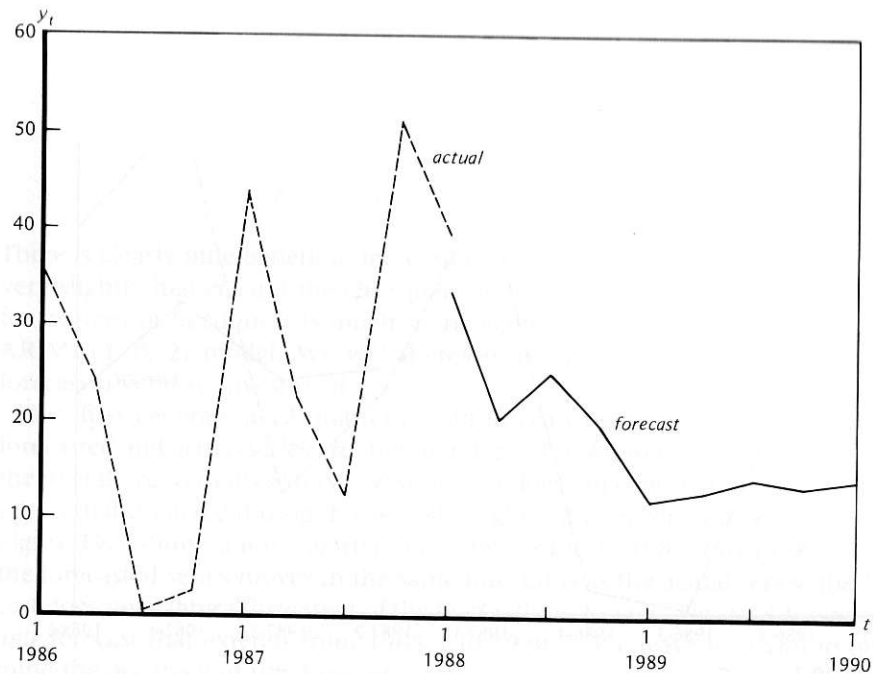
FIGURE 19.8
Ex ante forecast of inventory investment.

19.3 FORECASTING SEASONAL TELEPHONE DATA

An article by Thompson and Tiao provides another interesting case study of time-series analysis.[2] In the study forecasting models were constructed for the inward and outward station movements of the Wisconsin Telephone Company using monthly data from January 1951 to October 1966. The inward station movement in a given month is the sum of residence and business telephone installations, while the outward station movement consists of removals and disconnects of telephones. It is important to the telephone company to obtain reasonably accurate forecasts of station movements, since these forecasts are used as fundamental inputs to both short- and long-term company planning. The difference between inward and outward station movements represents the net increase (or decrease) of telephones in service, so that an expected positive difference would lead to a sequence of capital expenditures. Underestimating the difference might create a shortage in the supply of telephones and associated facilities, while overestimating it would result in a premature expansion of facilities and thus added cost to the company.

[2] H. E. Thompson and G. C. Tiao, "Analysis of Telephone Data: A Case Study of Forecasting Seasonal Time Series," *Bell Journal of Economics and Management Science*, vol. 2, no. 2, Autumn 1971.
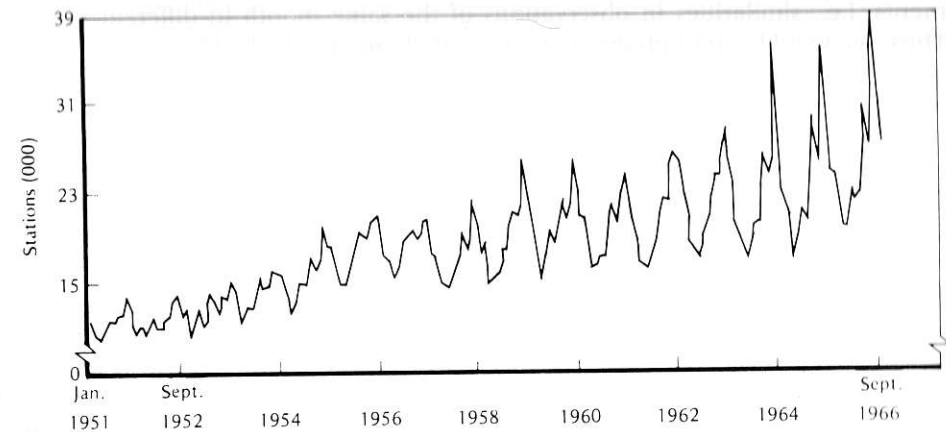
FIGURE 19.9
Monthly inward station movements, January 1951 to October 1966. (*Bell Journal of Economics and Management Sciences, vol. 2, no. 2, Autumn 1971.*)

The data used by Thompson and Tiao for inward and outward station movements are shown in Figs. 19.9 and 19.10. The data show a very distinct seasonal pattern, with a peak and a trough reached each year. Note that the *level* of each series tends to increase over time *and that the variance of the data tends to increase as the level increases*. In order to reduce this dependence of the variance on the level, the authors applied a logarithmic transformation to both series. Thus, the analysis that follows is given in terms of transformed logarithmic data. (Logarithmic transformations are often used in time-series analysis as a means of removing growth over time of the variance of the data.)

Time-series models can easily be constructed to account for seasonality; in fact, we treated seasonality earlier when we constructed a time-series model for hog production. It is reasonable to expect a seasonal pattern in station movements.

FIGURE 19.10
Monthly outward station movements, January 1951 to October 1966 (y_t). (*Bell Journal of Economics and Management Sciences, vol. 2, no. 2, Autumn 1971.*)
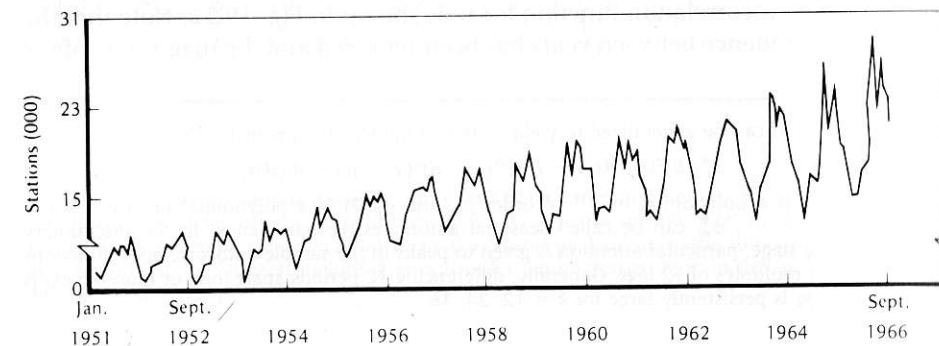
ments, i.e., similarities in observations of the same month in different years. Thus, we would expect observations 12 periods apart to be highly correlated (as in our hog production example). We can express this seasonal relationship with the simple autoregressive model

$$(1 - \phi^*B^{12})y_t = e_t \qquad (19.5)$$

where e_t is a random shock. While this equation explains observations between years, observations in successive months may also be dependent. This dependence might be represented by a second autoregressive model:

$$(1 - \phi B)e_t = \varepsilon_t \qquad (19.6)$$

where ε_t is a random shock. Equation (19.5) can be substituted into (19.6) to eliminate e_t:

$$(1 - \phi^*B^{12})(1 - \phi B)y_t = \varepsilon_t \qquad (19.7)$$

or

$$y_t - \phi y_{t-1} - \phi^*y_{t-12} + \phi\phi^*y_{t-13} = \varepsilon_t \qquad (19.8)$$

Equation (19.8) is a simple autoregressive model. It serves to describe, however, both seasonal and nonseasonal dependence between observations.[3]

In this case we present Thompson and Tiao's model of the *logarithmic outward series*. (The reader interested in the rest of their results may refer to the original paper.) We represent the logarithm of monthly outward station movements by the variable y_t. The sample autocorrelation function of y_t is shown in Fig. 19.11. Note that this autocorrelation function peaks at $k = 12$, 24, and 36, which is not surprising in view of the seasonal pattern in the data. We thus calculate 12-period differences in the series and call this new series w_t:

$$w_t = (1 - B^{12})y_t \qquad (19.9)$$

The sample autocorrelation function for w_t is shown in Fig. 19.12. Note that the seasonal dependence between years has been removed and the magnitude of the

---

[3] This equation can be generalized to yield a class of models for seasonal series:

$$\phi^*_{p_1}(B^{12})\phi_p(B)(1 - B^{12})^{d_1}(1 - B)^d(y_t - \mu) = \theta_q(B)\varepsilon_t$$

where $\phi^*_{p_1}(B^{12})$ is a polynomial in B^{12} of order p_1, and $\phi_p(B)$ is a polynomial of order p. The parameters $\phi^*_1, \ldots, \phi^*_{p_1}$ can be called seasonal autoregressive parameters. In the preliminary model-building stage, particular attention is given to peaks in the sample autocorrelation functions which occur at multiples of 12 lags. Generally, differencing 12 periods apart (one or more times) is needed when ρ_k is persistently large for $k = 12$, 24, 36, ...
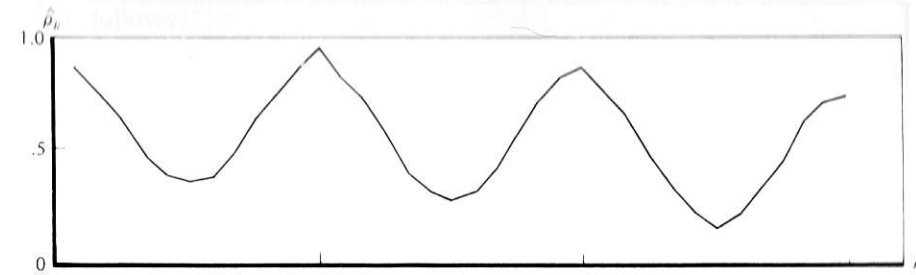

FIGURE 19.11
Sample autocorrelation function of y_t. (*Bell Journal of Economics and Management Sciences*, vol. 2, no. 2, Autumn 1971.)

autocorrelations has been dampened considerably. Also, note that this autocorrelation function has peaks at every third lag, thus suggesting the autoregressive model[4]

$$(1 - \phi_3 B^3)w_t = \varepsilon_t \qquad (19.10)$$

Thompson and Tiao fitted a third-order autoregressive model to the series w_t and then calculated the autocorrelation function for the residuals of this model. They found peaks at $k = 9$, 12, and 13, suggesting the addition of three moving average parameters. Thus, their final ARIMA model for y_t was of the form

$$(1 - \phi_3 B^3)(1 - \phi_{12}B^{12})y_t = (1 - \theta_9 B^9 - \theta_{12}B^{12} - \theta_{13}B^{13})\varepsilon_t \qquad (19.11)$$

FIGURE 19.12
Sample autocorrelation function for w_t. (*Bell Journal of Economics and Management Sciences*, vol. 2, no. 2, Autumn 1971.)

---
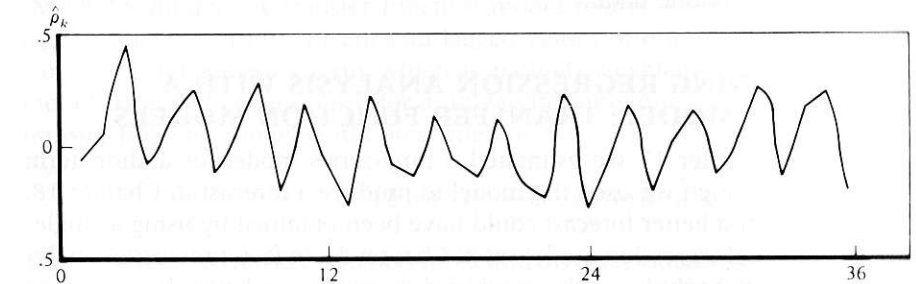
[4] Cycles every third period could also be generated by a second-order autoregressive model (with the proper parameter values). The authors may have tested a second-order model and found Eq. (19.10) to be preferable. In general, however, if a distinct peak occurs in the autocorrelation function at every nth lag, we suggest including an nth-order autoregressive term in the specification of the ARIMA model.
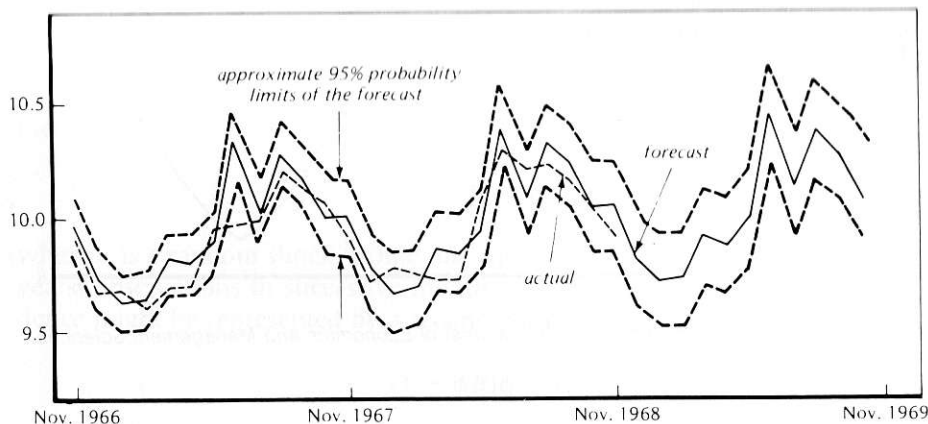
FIGURE 19.13
Forecasts of log outward series for the 36 months, November 1966 to October 1969, made in October 1966. (*Bell Journal of Economics and Management Sciences*, vol. 2, no. 2, Autumn 1971.)

The five parameters ϕ_3, ϕ_{12}, θ_9, θ_{12}, and θ_{13} were estimated, and the resulting model was used to forecast the logarithmic outward series for the 36 months from November 1966 to October 1969. The forecast, together with the 95 percent confidence interval, is shown in Fig. 19.13.

Note that the model does a rather good job of forecasting outward station movements, even over a period of 36 months. In fact, it seems to perform considerably better than our models of inventory investment did. The reason for this is that the telephone data used in Thompson and Tiao's study were particularly amenable to time-series analysis. Time-series analysis works best when a persistent pattern (seasonal or otherwise) exists in the data, and such a pattern is present in the telephone data.

19.4 COMBINING REGRESSION ANALYSIS WITH A TIME-SERIES MODEL: TRANSFER FUNCTION MODELS

At the end of Chapter 17 we estimated a time-series model for a short-term interest rate. Although we used the model to produce a forecast in Chapter 18, we suggested that a better forecast could have been obtained by using a single-equation structural regression model (as in Chapter 8). In fact, time-series analysis and regression analysis can be combined to produce a better forecast than would be possible through the use of either of these techniques alone.

Suppose that we would like to forecast the variable y_t using a regression model. Presumably such a model would include those independent variables which can explain movements in y_t but which are not themselves collinear. Suppose that our regression model contains two independent variables, x_1 and

x_2, as follows:

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t \qquad (19.12)$$

This equation has an additive error term that accounts for *unexplained* variance in y_t; that is, it accounts for that part of the variance of y_t that is not explained by x_1 and x_2. The equation can be estimated, and an R^2 will result which (unless by some chance y_t is perfectly correlated with the independent variables) will be less than 1. The equation can then be used to forecast y_t. As we saw in Chapter 8, one source of forecast error would come from the additive noise term whose future values cannot be predicted.

One effective application of time-series analysis is to construct an ARIMA model for the residual series u_t of this regression. We would then substitute the ARIMA model for the implicit error term in the original regression equation. When using the equation to forecast y_t, we would also be able to make a forecast of the error term ε_t using the ARIMA model. The ARIMA model provides some information as to what future values of ε_t are likely to be; i.e., it helps "explain" the unexplained variance in the regression equation. The combined regression–time-series model is

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \phi^{-1}(B)\theta(B)\eta_t \qquad (19.13)$$

where η_t is a normally distributed error term which may have a different variance from ε_t. This model is likely to provide better forecasts than the regression equation (19.12) alone or a time-series model alone since it includes a structural (economic) explanation of that part of the variance of y_t that can be explained structurally, and a time-series "explanation" of that part of the variance of y_t that cannot be explained structurally.

Equation (19.13) is an example of what is sometimes referred to as a *transfer function model* or, alternatively, a *multivariate autoregressive–moving average model* (MARMA model). A transfer function model relates a dependent variable to lagged values of itself, current and lagged values of one or more independent variables, and an error term which is partially "explained" by a time-series model. Thus the general form for a univariate (only one independent variable) transfer function model could be written as

$$y_t = \nu^{-1}(B)\omega(B)x_t + \phi^{-1}(B)\theta(B)\eta_t \qquad (19.14)$$

The technique of transfer function modeling involves examination of partial and total autocorrelation functions for the independent variable x_t as well as the dependent variable y_t in an effort to specify the lag polynomials $\nu(B)$, $\omega(B)$, $\phi(B)$, and $\theta(B)$.[5] One problem with the technique, however, is that the specifi-

[5] The techniques are discussed in detail in G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (San Francisco: Holden-Day, 1970), chaps. 10 and 11, and C. W. J. Granger and P. Newbold, op. cit.

cation of the structural part of the model, i.e., the polynomials $\nu(B)$ and $\omega(B)$, is done mechanically, rather than by appeal to economic theory and logic. Structural models that are consistent with intuition and economic theory are usually more reliable (and defensible) than models in which the structure is arrived at mechanically. For this reason we suggest that models of the form of Eq. (19.14) be used, but that the structural part of the model be arrived at through the mixture of economic theory and econometric method discussed in Part One, while the time-series part of the model, that is, $\phi(B)$ and $\theta(B)$, be arrived at through an analysis of the residuals of the structural model.

Let us now turn back to the simple model of Eq. (19.13). First, note that specifying a time-series model for the error term is just a generalization of the technique described in Chapter 8 for forecasting with regression models that have serially correlated errors. [If the time-series model is AR(1), it is exactly equivalent to forecasting with first-order serially correlated errors.] Second, note that the parameters a_0, a_1, and a_2 of the structural regression equation and the parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ of the time-series model should be estimated *simultaneously*. (Failure to estimate all the parameters simultaneously can lead to a loss of efficiency.) Unfortunately, the simultaneous estimation of all the parameters is sometimes computationally difficult and in such cases is not done.

This combined use of regression analysis with a time-series model of the error term is an approach to forecasting that in some cases can provide the best of both worlds. To demonstrate the technique and its use, we turn to two examples.

19.5 A COMBINED REGRESSION–TIME-SERIES MODEL TO FORECAST SHORT-TERM SAVINGS DEPOSIT FLOWS

Our first example that combines time-series analysis with regression analysis is based on a study by Ludwig[6] to forecast the monthly flow of deposits into Massachusetts mutual savings banks. A regression model is first constructed (to explain deposit flows), and then a time-series model is developed to "explain" the residual series (i.e., the error term) in the regression equation.[7]

We begin with a regression equation that provides a structural explanation of mutual savings deposit flows. Ludwig used the ratio of deposit flows S to personal wealth W as the dependent variable, and he chose monthly Massachusetts personal income as a proxy variable for wealth. His best regression equation had three explanatory variables: the effective percentage return (including dividends) on mutual savings deposits r_{ms}, the interest rate on 3-month Treasury bills r_m, and the ratio of the previous month's stock of mutual savings deposits

[6] R. S. Ludwig, "Forecasting Short-Term Savings Deposit Flows: An Application of Time Series Models and a Regional Analysis," unpublished Master's thesis, Sloan School of Management, M.I.T., June 1974.

[7] At the time this study was done, simultaneous estimation of the regression and time-series parameters was computationally difficult and so was not performed.

A_{-1} to the wealth variable. His equation, estimated using monthly data for the state of Massachusetts over the period February 1968 to June 1973, is

$$\frac{S}{W} = \underset{(1.89)}{.16} + \underset{(2.98)}{.019} r_{ms} - \underset{(-5.27)}{.011} r_m - \underset{(-2.23)}{.032} \frac{A_{-1}}{W} \tag{19.15}$$

$$R^2 = .41 \qquad SER = .016 \qquad F = 14.42 \qquad DW = 1.55$$

As one would expect, there is a positive relationship between savings deposit flows and the effective percentage return on deposits. The interest rate on 3-month Teasury bills, used as a market rate of interest, represents the return on competing risk-free investment alternatives for savings, and thus should have a negative impact on savings deposit flows. Finally, the negative relationship between deposit flows and the stock of deposits represents a stock adjustment effect; savings deposits should be proportional to that part of personal wealth that has not already been placed in a savings bank; i.e.,
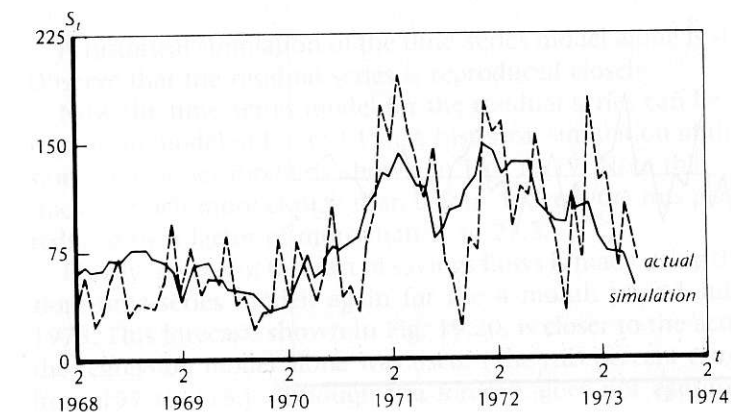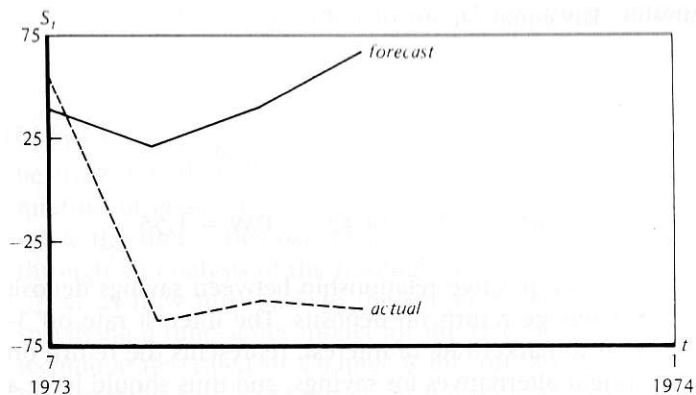
$$S_t = A_t - A_{t-1} = a(W_t - A_{t-1}) \tag{19.16}$$

so that

$$\frac{S_t}{W_t} = a - a\frac{A_{t-1}}{W_t} \tag{19.17}$$

A historical simulation of Eq. (19.15) is shown in Fig. 19.14, and an *ex post* forecast over the period July 1973 to October 1973 is shown in Fig. 19.15. The historical simulation has an rms *percent* error of 75.1, and the *ex post* forecast has an rms percent error of 157. Observe that the simulation tracks the general movement of the series but leaves much of the variance unexplained. The re-

FIGURE 19.14
Historical simulation of Eq. (19.15) for deposit flows.

FIGURE 19.15
Ex post forecast of Eq. (19.15) for deposit flows.

FIGURE 19.17
Autocorrelation function of 12-month difference of residuals $(1 - B^{12})u_t$.

gression model does well in forecasting deposit flows in July 1973 but fails to capture the sharp drop in deposits in August of that year.

Let us now see how to improve the forecast by constructing a time-series model for the residual series of the regression equation. The sample autocorrelation function for the residual series is shown in Fig. 19.16. Observe that high-order correlations damp toward 0, so that the residual series can be considered stationary. The autocorrelation function does, however, contain peaks at monthly lags which are multiples of 12, indicating annual seasonality. Figure 19.17 shows the sample autocorrelation function for a 12-month difference of the original residual series, i.e., for the series $(1 - B^{12})u_t$. This autocorrelation function has a damped sinusoidal shape which is indicative of a purely autoregressive process of order 2 or greater.

FIGURE 19.16
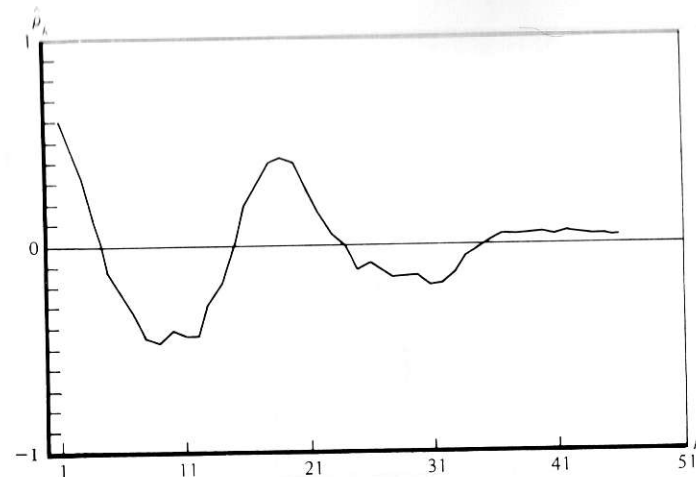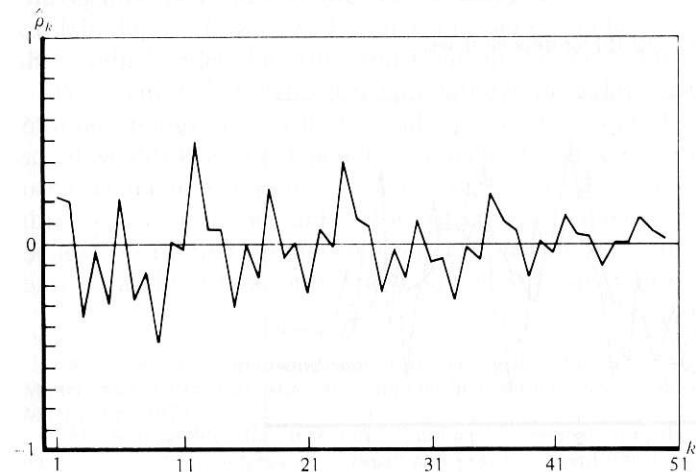Autocorrelation function of residuals u_t from Eq. (19.15).

Ludwig estimated a variety of autoregressive models for this residual series and found the best model to be of the form

$$(1 - \phi_{12}B^{12})(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6)u_t = \eta_t$$

(19.18)

which in its expanded and estimated form is

$$(1 - .736B - .025B^2 - .055B^3 - .009B^4 + .310B^5 - .128B^6 - .782B^{12}$$
$$+ .532B^{13} + .081B^{14} + .125B^{15} - .213B^{16} - .103B^{17} - .060B^{18})u_t = \eta_t$$

(19.19)

$$R^2 = .78 \qquad \chi^2 = 14.5$$

A historical simulation of the time-series model alone is shown in Fig. 19.18. Observe that the residual series is reproduced closely.

Now the time-series model for the residual series can be combined with the regression model of Eq. (19.15). A historical simulation of the combined regression–time-series model is shown in Fig. 19.19. Note that savings deposits are tracked much more closely than before. Indeed, the rms percent error has been reduced by a factor of more than 3, to 29.3.

Finally, an *ex post* forecast of savings flows is made using the combined regression–time-series model, again for the 4-month period July 1973 to October 1973. This forecast, shown in Fig. 19.20, is closer to the actual data than when the regression model alone was used. (The rms percent error has been reduced from 157 to 118.) Although the forecast does not capture the extent of the
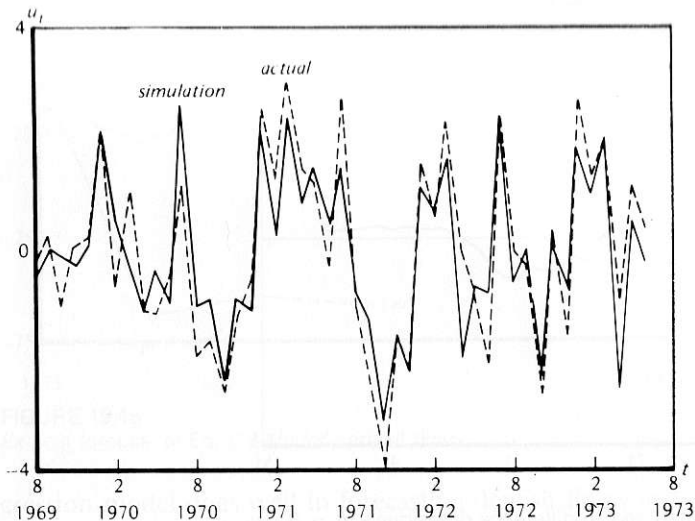
FIGURE 19.18
Historical simulation of time-series model for residuals.

downturn in savings deposit flows in August 1973, it does capture general movements in the variable.

19.6   A COMBINED REGRESSION–TIME-SERIES MODEL TO FORECAST INTEREST RATES

As a second example of the combined use of regression analysis with time-series models, we construct a model to forecast, on a monthly basis, the interest rate on

FIGURE 19.19
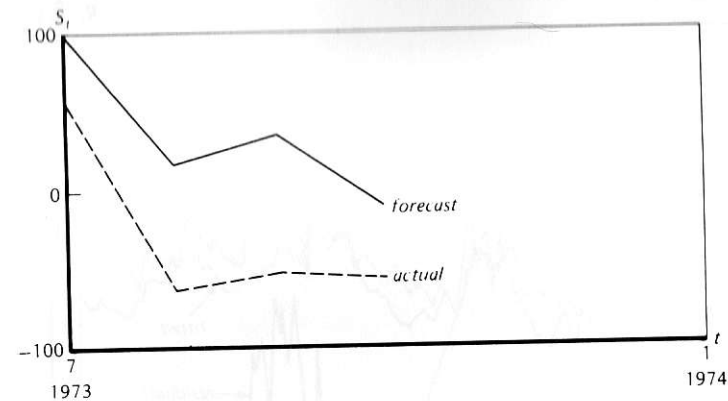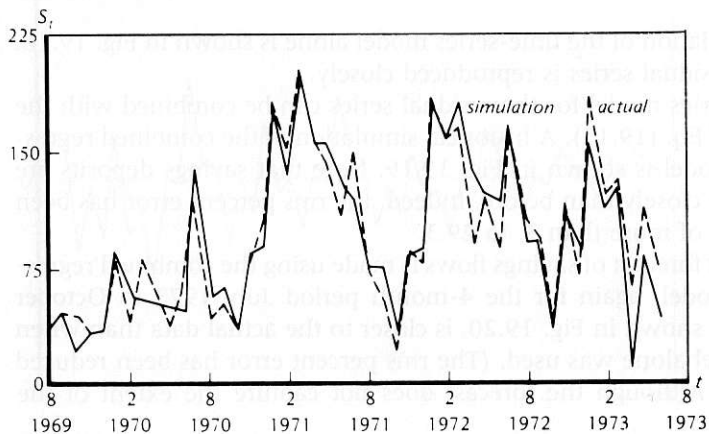Historical simulation of combined regression–time-series model for savings deposit flows.

FIGURE 19.20
Ex post forecast of savings deposit flows using combined regression–time-series model.

3-month Treasury bills. We will begin with a simple regression model that explains the interest rate as a function of industrial production, inflation, and the rate of growth of the money supply, all lagged. We will then examine the residuals of that model and fit an ARIMA model to them. Finally, we will reestimate all the parameters of the combined regression–time-series model simultaneously (a step that we did not take in the previous example).[8]

We use the following notation: R is the Treasury bill rate (in percent per annum), IP is the Index of Industrial Production, GM is the monthly percentage rate of growth of the (narrowly defined) money supply, and INF is the monthly percentage rate of growth of the Producer Price Index. We estimated our regression model using monthly data from January 1960 to March 1988 and obtained (t statistics in parentheses):

$$R = -.8010 - 21.924\text{GM}_{-1} + .0749\text{IP}_{-1} - 24.608(\Delta\text{IP}_{-1}/\text{IP}_{-2})$$
$$\quad (-1.86) \quad (-.93) \quad\quad (14.99) \quad\quad (-2.11)$$

$$\quad + 31.968\text{INF}_{-1} + 39.730\text{INF}_{-2} + 45.475\text{INF}_{-3} + 39.828\text{INF}_{-4} \quad\quad (19.20)$$
$$\quad\quad (1.86) \quad\quad\quad (2.32) \quad\quad\quad (2.65) \quad\quad\quad (2.29)$$

$$R^2 = .528 \quad\quad s = 2.036 \quad\quad F = 53.36 \quad\quad \text{DW} = .125$$

Figure 19.21 shows the actual and fitted series, together with the regression residuals. Note that the residuals appear to have a high degree of positive autocorrelation, which is consistent with the very low Durbin-Watson statistic. Also

[8] In the first and second editions of this book we wrote that "Unfortunately, the simultaneous estimation of all the parameters can sometimes entail considerable computational expense, and therefore is often not done in practice." The cost of computing has fallen dramatically since the publication of our second edition in 1981, and better software has also become available. The estimations in this example were done quite easily using MicroTSP and can also be done easily using the micro version of SHAZAM, as well as other microcomputer and mainframe packages.
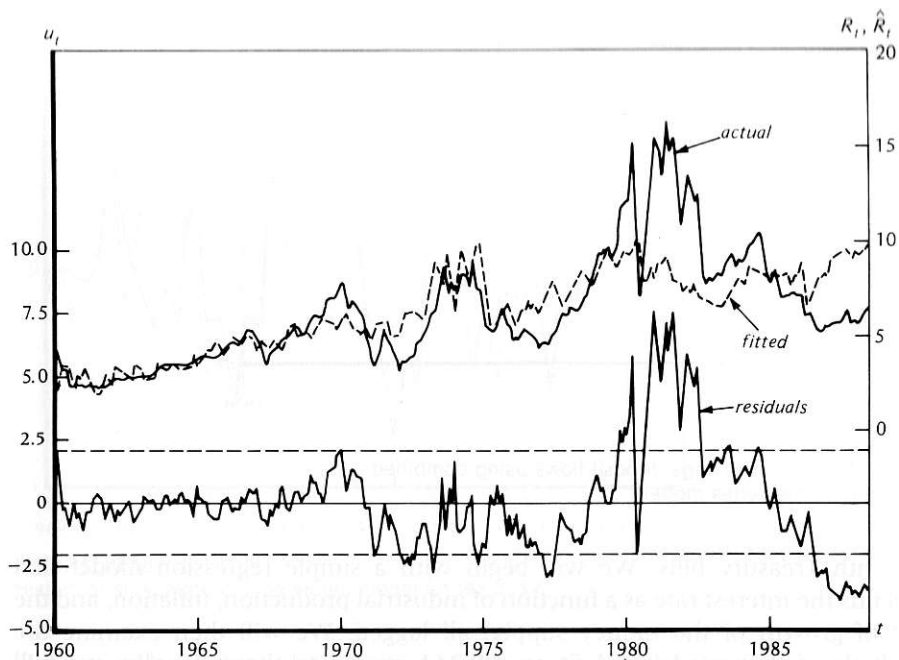
FIGURE 19.21
Three-month Treasury bill rate, actual, fitted, and residuals.

FIGURE 19.22
Regression model: 3-month forecast of Treasury bill rate.

FIGURE 19.23
Regression residuals: sample autocorrelation function.

note that the model fits the data reasonably well during the 1960s, less well during the 1970s, and rather poorly during the 1980s. This is not surprising, given that interest rates became extremely volatile beginning in the late 1970s (in part because of a change in Federal Reserve operating policy). In fact, most large econometric models failed miserably to predict interest rate changes during this period. Our regression model does no better in predicting rates. Figure 19.22 shows a 3-month *ex post* forecast for April, May, and June of 1988; the forecast is clearly way off the mark.

Let us now examine the residuals from this regression. Figure 19.23 shows the sample autocorrelation function for the residuals, which declines steadily toward zero, indicative of a stationary series. Figure 19.24 shows the sample autocorrelation function for the residuals after they have been first-differenced; all the autocorrelations are close to zero. We will work with the undifferenced residuals.

After some experimenting, we fit the following ARIMA(12, 0, 2) model to the series of residuals, which we denote by u_t:

$$(1 - 1.1089B + .3157B^2 - .1159B^3 + .1886B^4 - .4076B^5$$
$$+ .3693B^6 - .1075B^7 - .0635B^8 - .1786B^9 + .2202B^{10} - .0850B^{11}$$
$$+ .0038B^{12})u_t = -.3528 + (1 - .0007B + .0025B^2)\eta_t \qquad (19.21)$$
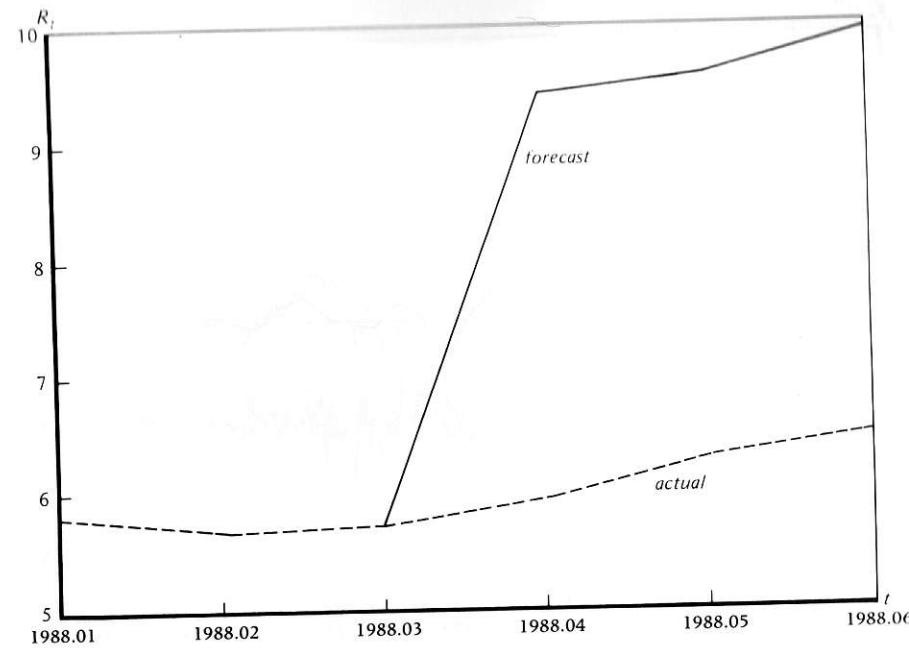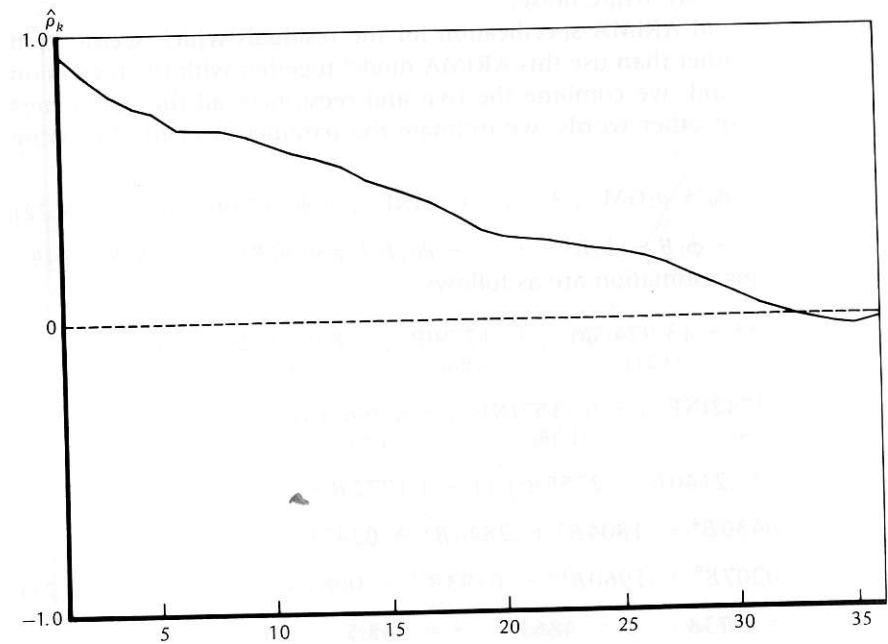$$R^2 = .900 \qquad \chi^2(14, 36) = 25.27$$

FIGURE 19.24
First differences of regression residuals: sample autocorrelation function.

FIGURE 19.25
Three-month Treasury bill rate, actual, fitted, and residuals (combined regression–time-series model).

FIGURE 19.26
Combined regression–time-series model: 3-month *ex post* forecast.

With $36 - 14 = 22$ degrees of freedom, the chi-square statistic is insignificant at the 90 percent level, so that we can accept the hypothesis that the residuals of this ARIMA model are white noise.

We now have an ARIMA specification for the residuals which seems to fit well. However, rather than use this ARIMA model together with the regression model as they stand, we combine the two and reestimate all the parameters simultaneously. In other words, we estimate the parameters of the following model:

$$R_t = a_0 + a_1 GM_{t-1} + \ldots + a_7 INF_{t-4} + \phi^{-1}(B)\theta(B)\eta_t \qquad (19.22)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_{12}B^{12}$, and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$. The results of this estimation are as follows:

$$R_t = -17.735 + 13.974 GM_{t-1} + .1779 IP_{t-1} - 8.0316(\Delta IP_{t-1}/IP_{t-2})$$
$$\quad\;\;(-.83) \qquad\quad (3.21) \qquad\qquad (3.96) \qquad\qquad (-2.59)$$

$$+\; 13.9742 INF_{t-1} + 6.7157 INF_{t-2} + 6.2660 INF_{t-3} - 3.3522 INF_{t-4}$$
$$\qquad (2.38) \qquad\qquad (1.38) \qquad\qquad (1.28) \qquad\qquad (-.87)$$

$$+\; \{(1 + .2140B - .2755B^2)/(1 - 1.1772B + .1138B^2 + .1876B^3$$

$$\qquad - .0430B^4 - .1804B^5 + .2846B^6 - .0247B^7 - .2844B^8$$

$$\qquad - .0207B^9 + .1960B^{10} - .0393B^{11} - .0081B^{12})\}\eta_t \qquad (19.23)$$

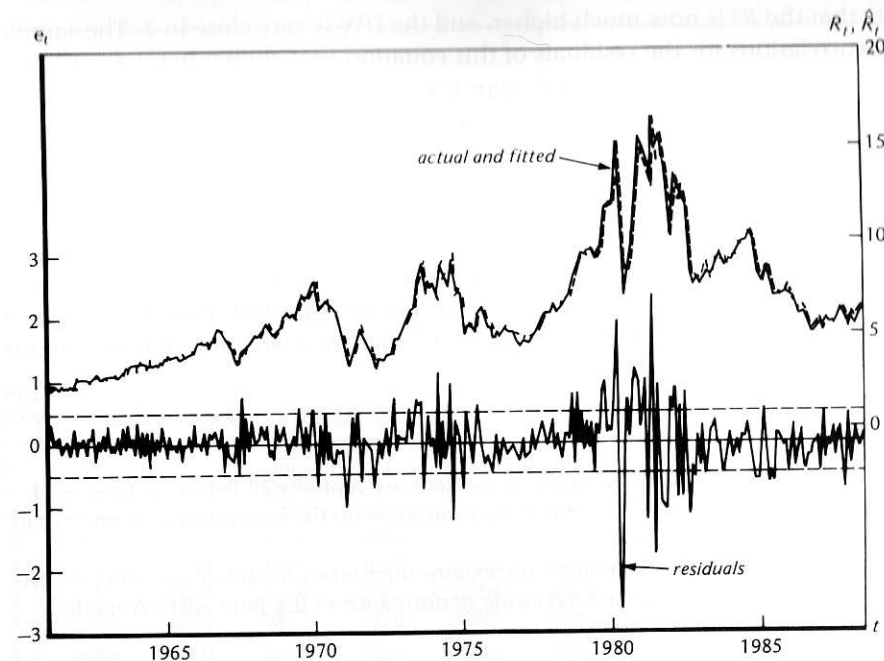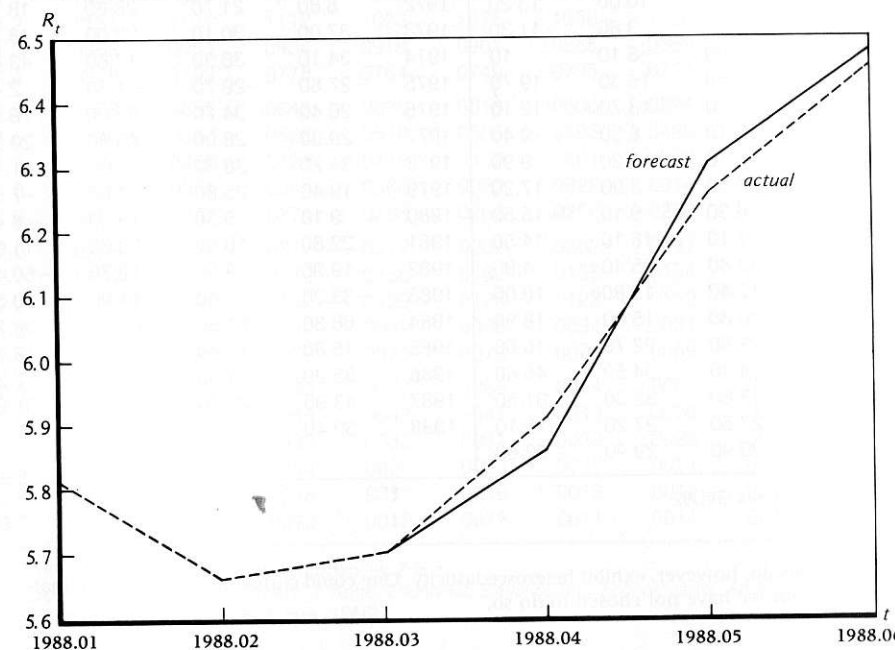$$R^2 = .9738 \qquad s = .4863 \qquad F = 545.5 \qquad DW = 2.001$$

Note that the R^2 is now much higher, and the DW is very close to 2. The sample autocorrelations for the residuals of this equation (not shown here) are all very close to zero, so that the residuals appear to be white noise. Figure 19.25 shows the fitted and actual interest rate series, as well as the residuals. Unlike the simple regression model that we started with, the fit of this equation is excellent throughout the sample period, and the residuals exhibit no autocorrelation.[9]

Finally, Fig. 19.26 shows an *ex post* forecast for the last three months of the sample period. Note that the forecasted values of the interest rate are now quite close to the actual values. This combined regression–time-series model seems to be a much better forecasting tool than the simple regression model, and it also forecasts better than the pure ARIMA model that we developed and used in our earlier interest rate examples.

EXERCISE

19.1 The data for nonfarm inventory investment are reproduced below in Table 19.1.

(a) Try to develop an ARIMA model that improves on the forecasting performance of the one presented in Section 19.2.

(b) Develop your own combined regression–time-series model of inventory investment. Can you improve on the forecasting performance of the pure ARIMA model?

TABLE 19.1
REAL NONFARM INVENTORY INVESTMENT

Obs.					Obs.				
1950	10.00	14.70	13.40	45.40	1970	1.30	8.50	16.50	3.30
1951	26.40	41.20	28.40	12.10	1971	26.00	16.10	16.80	5.50
1952	13.10	−8.70	10.00	13.20	1972	8.80	21.70	28.60	18.10
1953	8.70	11.20	3.80	−11.30	1973	37.00	30.10	22.00	48.70
1954	−7.60	−10.00	−8.10	.10	1974	34.10	38.30	17.60	43.60
1955	11.30	17.50	16.30	19.70	1975	−27.80	−29.70	−1.30	2.40
1956	18.90	16.20	13.70	12.10	1976	26.40	34.70	25.00	16.90
1957	5.50	5.10	6.30	−9.40	1977	29.90	28.00	45.90	20.00
1958	−16.90	−15.10	−1.20	9.90	1978	34.20	38.30	31.00	41.10
1959	14.60	27.50	3.90	17.20	1979	19.40	25.80	4.00	−7.50
1960	25.50	6.30	9.10	−15.50	1980	9.10	9.30	−19.00	−8.40
1961	−8.00	1.10	16.10	14.50	1981	22.80	10.60	30.60	11.90
1962	19.50	13.80	15.40	4.90	1982	−19.90	−9.50	−12.70	−50.40
1963	13.90	15.40	17.80	10.00	1983	−33.20	−2.40	14.80	20.50
1964	18.60	16.80	15.90	18.90	1984	68.30	61.50	62.60	38.70
1965	31.00	21.50	22.70	16.60	1985	15.80	12.40	3.20	16.70
1966	33.60	34.40	34.50	45.60	1986	35.20	23.90	.10	2.30
1967	31.40	13.80	32.30	31.50	1987	43.90	22.70	12.10	51.50
1968	11.90	27.50	27.20	15.10	1988	39.40			
1969	24.10	20.90	29.90	22.60					

Source: Citibase, Series GVU82.

[9] The residuals do, however, exhibit heteroscedasticity. One could correct for this when estimating the model, but we have not chosen to do so.

TABLE 1
STANDARDIZED NORMAL DISTRIBUTION

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
.3	.3821	.3873	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2217	.2148
.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0366	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0020	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0010	.0011	.0010

The table plots the cumulative probability $Z \geq z$.

Source: Produced from Edward J. Kane, *Economic Statistics and Econometrics: An Introduction to Quantitative Economics* (New York: Harper & Row, 1968).