CHAPTER 14 Introduction to Time Series Regression and Forecasting

Time series data—data collected for a single entity at multiple points in time—can be used to answer quantitative questions for which cross-sectional data are inadequate. One such question is, what is the causal effect on a variable of interest, *Y*, of a change in another variable, *X*, over time? In other words, what is the *dynamic* causal effect on *Y* of a change in *X*? For example, what is the effect on traffic fatalities of a law requiring passengers to wear seatbelts, both initially and subsequently as drivers adjust to the law? Another such question is, what is your best forecast of the value of some variable at a future date? For example, what is your best forecast of next month's rate of inflation, interest rates, or stock prices? Both of these questions—one about dynamic causal effects, the other about economic forecasting—can be answered using time series data. But time series data pose special challenges, and overcoming those challenges requires some new techniques.

Chapters 14 through 16 introduce techniques for the econometric analysis of time series data and apply these techniques to the problems of forecasting and estimating dynamic causal effects. Chapter 14 introduces the basic concepts and tools of regression with time series data and applies them to economic forecasting. In Chapter 15, the concepts and tools developed in Chapter 14 are applied to the problem of estimating dynamic causal effects using time series data. Chapter 16 takes up some more advanced topics in time series analysis, including forecasting multiple time series and modeling changes in volatility over time.

The empirical problem studied in this chapter is forecasting the rate of inflation, that is, the percentage increase in overall prices. While in a sense forecasting is just an application of regression analysis, forecasting is quite different from the estimation of causal effects, the focus of this book until now. As discussed in Section 14.1, models that are useful for forecasting need not have a causal interpretation: If you see pedestrians carrying umbrellas you might forecast rain, even though carrying an umbrella does not *cause* it to rain. Section 14.2 introduces some basic concepts of time series analysis and presents some examples of economic time series data. Section 14.3 presents time series regression models in which the regressors are past values of the dependent variable; these "autoregressive" models use the history of inflation to forecast its future. Often, forecasts based on autoregressions can be improved by adding additional predictor variables and their past values, or "lags," as regressors, and these so-called autoregressive distributed lag models are introduced in Section 14.4. For example, we find that inflation forecasts made using lagged values of the rate of unemployment in addition to lagged inflation—that is, forecasts based on an empirical Phillips curve—improve upon the autoregressive inflation forecasts. A practical issue is deciding how many past values to include in autoregressions and autoregressive distributed lag models, and Section 14.5 describes methods for making this decision.

The assumption that the future will be like the past is an important one in time series regression, sufficiently so that it is given its own name, "stationarity." Time series variables can fail to be stationary in various ways, but two are especially relevant for regression analysis of economic time series data: (1) the series can have persistent, long-run movements, that is, the series can have trends; and (2) the population regression can be unstable over time, that is, the population regression can have breaks. These departures from stationarity jeopardize forecasts and inferences based on time series regression. Fortunately, there are statistical procedures for detecting trends and breaks and, once detected, for adjusting the model specification. These procedures are presented in Sections 14.6 and 14.7.

14.1 Using Regression Models for Forecasting

The empirical application of Chapters 4 through 9 focused on estimating the causal effect on test scores of the student–teacher ratio. The simplest regression model in Chapter 4 related test scores to the student–teacher ratio (*STR*):

$$\widehat{TestScore} = 989.9 - 2.28 \times STR. \tag{14.1}$$

As was discussed in Chapter 6, a school superintendent, contemplating hiring more teachers to reduce class sizes, would not consider this equation to be very helpful. The estimated slope coefficient in Equation (14.1) fails to provide a useful estimate of the causal effect on test scores of the student–teacher ratio because of probable omitted variable bias arising from the omission of school and student characteristics that are determinants of test scores and that are correlated with the student–teacher ratio.

In contrast, as was discussed in Chapter 9, a parent who is considering moving to a school district might find Equation (14.1) more helpful. Even though the coefficient does not have a causal interpretation, the regression could help the parent forecast test scores in a district for which they are not publicly available. More generally, a regression model can be useful for forecasting even if none of its coefficients has a causal interpretation. From the perspective of forecasting, what is important is that the model provides as accurate a forecast as possible. Although there is no such thing as a perfect forecast, regression models can nevertheless provide forecasts that are accurate and reliable.

The applications in this chapter differ from the test score/class size prediction problem because this chapter focuses on using time series data to forecast future events. For example, the parent actually would be interested in test scores next year, after his or her child has enrolled in a school. Of course, those tests have not yet been given, so the parent must forecast the scores using currently available information. If test scores are available for past years, then a good starting point is to use data on current and past test scores to forecast future test scores. This reasoning leads directly to the autoregressive models presented in Section 14.3, in which past values of a variable are used in a linear regression to forecast future values of the series. The next step, which is taken in Section 14.4, is to extend these models to include additional predictor variables such as data on class size. Like Equation (14.1), such a regression model can produce accurate and reliable forecasts even if its coefficients have no causal interpretation. In Chapter 15, we return to problems like that faced by the school superintendent and discuss the estimation of causal effects using time series variables.

14.2 Introduction to Time Series Data and Serial Correlation

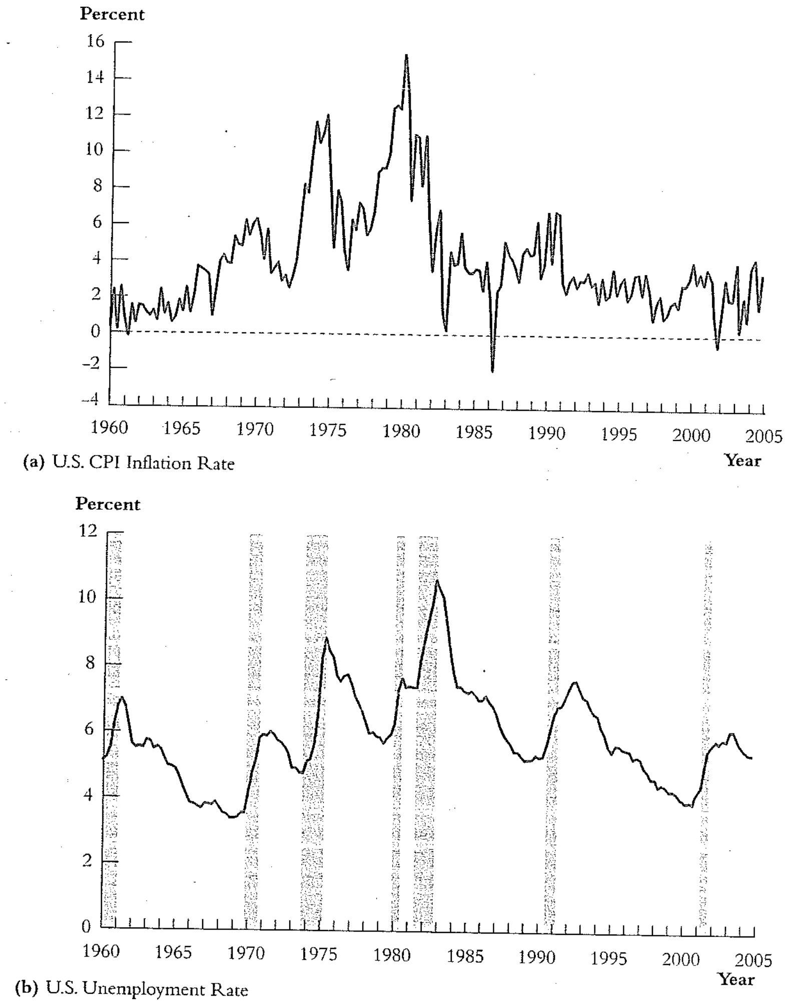
This section introduces some basic concepts and terminology that arise in time series econometrics. A good place to start any analysis of time series data is by plotting the data, so that is where we begin.

The Rates of Inflation and Unemployment in the United States

Figure 14.1a plots the U.S. rate of inflation—the annual percentage change in prices in the United States, as measured by the Consumer Price Index (CPI)—from 1960 to 2004 (the data are described in Appendix 14.1). The inflation rate was low in the 1960s, rose through the 1970s to a post–World War II peak of 15.5% in the first quarter of 1980 (that is, January, February, and March 1980), and then fell to less than 3% by the end of the 1990s. As can be seen in Figure 14.1a, the inflation rate also can fluctuate by one percentage point or more from one quarter to the next.

The U.S. unemployment rate—the fraction of the labor force out of work, as measured in the Current Population Survey (see Appendix 3.1)—is plotted in

FIGURE 14.1 Inflation and Unemployment in the United States, 1960–2004



(a) U.S. CPI Inflation Rate

(b) U.S. Unemployment Rate

Price inflation in the United States (Figure 14.1a) drifted upward from 1960 until 1980 and then fell sharply during the early 1980s. The unemployment rate in the United States (Figure 14.1b) rises during recessions (the shaded episodes) and falls during expansions.

Figure 14.1b. Changes in the unemployment rate are mainly associated with the business cycle in the United States. For example, the unemployment rate increased during the recessions of 1960–1961, 1970, 1974–1975, the twin recessions of 1980 and 1981–1982, and the recessions of 1990–1991 and 2001, episodes denoted by shading in Figure 14.1b.

Lags, First Differences, Logarithms, and Growth Rates

The observation on the time series variable Y made at date t is denoted Y_t, and the total number of observations is denoted T. The interval between observations, that is, the period of time between observation t and observation $t + 1$, is some unit of time such as weeks, months, quarters (three-month units), or years. For example, the inflation data studied in this chapter are quarterly, so the unit of time (a "period") is a quarter of a year.

Special terminology and notation are used to indicate future and past values of Y. The value of Y in the previous period is called its *first lagged value* or, more simply, its **first lag**, and is denoted Y_{t-1}. Its j^{th} *lagged value* (or simply its $\boldsymbol{j^{th}}$ **lag**) is its value j periods ago, which is Y_{t-j}. Similarly, Y_{t+1} denotes the value of Y one period into the future.

The change in the value of Y between period $t - 1$ and period t is $Y_t - Y_{t-1}$; this change is called the **first difference** in the variable Y_t. In time series data, "Δ" is used to represent the first difference, so $\Delta Y_t = Y_t - Y_{t-1}$.

Economic time series are often analyzed after computing their logarithms or the changes in their logarithms. One reason for this is that many economic series, such as gross domestic product (GDP), exhibit growth that is approximately exponential, that is, over the long run the series tends to grow by a certain percentage per year on average; if so, the logarithm of the series grows approximately linearly. Another reason is that the standard deviation of many economic time series is approximately proportional to its level, that is, the standard deviation is well expressed as a percentage of the level of the series; if so, then the standard deviation of the logarithm of the series is approximately constant. In either case, it is useful to transform the series so that changes in the transformed series are proportional (or percentage) changes in the original series, and this is achieved by taking the logarithm of the series.[1]

[1] The change of the logarithm of a variable is approximately equal to the proportional change of that variable; that is, $\ln(X + a) - \ln(X) \cong a/X$, where the approximation works best when a/X is small [see Equation (8.16) and the surrounding discussion]. Now, replace X with Y_{t-1}, and a with ΔY_t, and note that $Y_t = Y_{t-1} + \Delta Y_t$. This means that the proportional change in the series Y_t between periods $t - 1$ and t is approximately $\ln(Y_t) - \ln(Y_{t-1}) = \ln(Y_{t-1} + \Delta Y_t) - \ln(Y_{t-1}) \cong \Delta Y_t / Y_{t-1}$. The expression $\ln(Y_t) - \ln(Y_{t-1})$ is the first difference of $\ln(Y_t)$, $\Delta\ln(Y_t)$. Thus $\Delta\ln(Y_t) \cong \Delta Y_t / Y_{t-1}$. The percentage change is 100 times the fractional change, so the percentage change in the series Y_t is approximately $100\Delta\ln(Y_t)$.

Lags, First Differences, Logarithms, and Growth Rates

KEY CONCEPT

14.1

- The first lag of a time series Y_t is Y_{t-1}; its j^{th} lag is Y_{t-j}.
- The first difference of a series, ΔY_t, is its change between periods $t - 1$ and t; that is, $\Delta Y_t = Y_t - Y_{t-1}$.
- The first difference of the logarithm of Y_t is $\Delta\ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1})$.
- The percentage change of a time series Y_t between periods $t - 1$ and t is approximately $100\Delta\ln(Y_t)$, where the approximation is most accurate when the percentage change is small.

Lags, first differences, and growth rates are summarized in Key Concept 14.1.

Lags, changes, and percentage changes are illustrated using the U.S. inflation rate in Table 14.1. The first column shows the date, or period, where the first quarter of 2004 is denoted 2004:I, the second quarter of 2004 is denoted 2004:II, and so forth. The second column shows the value of the CPI in that quarter, and the third column shows the rate of inflation. For example, from the first to the second quarter of 2004, the index increased from 186.57 to 188.60, a percentage increase of $100 \times (188.60 - 186.57)/186.57 = 1.09\%$. This is the percentage increase from one

TABLE 14.1 Inflation in the United States in 2004 and the First Quarter of 2005

Quarter	U.S. CPI	Rate of Inflation at an Annual Rate (Inf_t)	First Lag (Inf_{t-1})	Change in Inflation (ΔInf_t)
2004:I	186.57	3.8	0.9	2.9
2004:II	188.60	4.4	3.8	0.6
2004:III	189.37	1.6	4.4	−2.8
2004:IV	191.03	3.5	1.6	1.9
2005:I	192.17	2.4	3.5	−1.1

The annualized rate of inflation is the percentage change in the CPI from the previous quarter to the current quarter, multiplied by four. The first lag of inflation is its value in the previous quarter, and the change in inflation is the current inflation rate minus its first lag. All entries are rounded to the nearest decimal.

quarter to the next. It is conventional to report rates of inflation (and other growth rates in macroeconomic time series) on an annual basis, which is the percentage increase in prices that would occur over a year, if the series were to continue to increase at the same rate. Because there are four quarters a year, the annualized rate of inflation in 2004:II is 1.09 × 4 = 4.36, or 4.4% per year after rounding.

This percentage change can also be computed using the differences-of-logarithms approximation in Key Concept 14.1. The difference in the logarithm of the CPI from 2004:I to 2004:II is ln(188.60) − ln(186.57) = 0.0108, yielding the approximate quarterly percentage difference 100 × 0.0108 = 1.08%. On an annualized basis, this is 1.08 × 4 = 4.32, or 4.3% after rounding, essentially the same as obtained by directly computing the percentage growth. These calculations can be summarized as

$$\begin{aligned}\text{Annualized rate of inflation} = Inf_t &\cong 400[\ln(CPI_t) - \ln(CPI_{t-1})] \\ &= 400\Delta\ln(CPI_t),\end{aligned} \tag{14.2}$$

where CPI_t is the value of the Consumer Price Index at date t. The factor of 400 arises from converting fractional change to percentages (multiplying by 100) and converting quarterly percentage change to an equivalent annual rate (multiplying by 4).

The final two columns of Table 14.1 illustrate lags and changes. The first lag of inflation in 2004:II is 3.8%, the inflation rate in 2004:I. The change in the rate of inflation from 2004:I to 2004:II was 4.4% − 3.8% = 0.6%.

KEY CONCEPT

14.2

Autocorrelation (Serial Correlation) and Autocovariance

The j^{th} autocovariance of a series Y_t is the covariance between Y_t and its j^{th} lag, Y_{t-j}, and the j^{th} autocorrelation coefficient is the correlation between Y_t and Y_{t-j}. That is,

$$j^{\text{th}} \text{ autocovariance} = \text{cov}(Y_t, Y_{t-j}) \tag{14.3}$$

$$j^{\text{th}} \text{ autocorrelation} = \rho_j = \text{corr}(Y_t, Y_{t-j}) = \frac{\text{cov}(Y_t, Y_{t-j})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-j})}}. \tag{14.4}$$

The j^{th} autocorrelation coefficient is sometimes called the j^{th} serial correlation coefficient.

Autocorrelation

In time series data, the value of Y in one period typically is correlated with its value in the next period. The correlation of a series with its own lagged values is called **autocorrelation** or **serial correlation.** The first autocorrelation (or **autocorrelation coefficient**) is the correlation between Y_t and Y_{t-1}, that is, the correlation between values of Y at two adjacent dates. The second autocorrelation is the correlation between Y_t and Y_{t-2}, and the j^{th} autocorrelation is the correlation between Y_t and Y_{t-j}. Similarly, the **j^{th} autocovariance** is the covariance between Y_t and Y_{t-j}. Autocorrelation and autocovariance are summarized in Key Concept 14.2.

The j^{th} population autocovariances and autocorrelations in Key Concept 14.2 can be estimated by the j^{th} sample autocovariances and autocorrelations, $\widehat{\text{cov}(Y_t, Y_{t-j})}$ and $\hat{\rho}_j$:

$$\widehat{\text{cov}(Y_t, Y_{t-j})} = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \overline{Y}_{j+1,T})(Y_{t-j} - \overline{Y}_{1,T-j}) \tag{14.5}$$

$$\hat{\rho}_j = \frac{\widehat{\text{cov}(Y_t, Y_{t-j})}}{\widehat{\text{var}(Y_t)}}, \tag{14.6}$$

where $\overline{Y}_{j+1,T}$ denotes the sample average of Y_t computed over the observations $t = j + 1, \ldots, T$ and where $\widehat{\text{var}(Y_t)}$ is the sample variance of Y.[2]

The first four sample autocorrelations of the inflation rate and of the change in the inflation rate are listed in Table 14.2. These entries show that inflation is strongly positively autocorrelated: The first autocorrelation is 0.84. The sample autocorrelation declines as the lag increases, but it remains large even at a lag of four quarters. The change in inflation is negatively autocorrelated: An increase in the rate of inflation in one quarter tends to be associated with a decrease in the next quarter.

At first, it might seem contradictory that the level of inflation is strongly positively correlated but its change is negatively correlated. These two autocorrelations, however, measure different things. The strong positive autocorrelation in inflation reflects the long-term trends in inflation evident in Figure 14.1: Inflation was low in the first quarter of 1965 and again in the second; it was high in the first quarter of 1981 and again in the second. In contrast, the negative autocorrelation of the change of inflation means that, on average, an increase in inflation in one quarter is associated with a decrease in inflation in the next.

[2] The summation in Equation (14.5) is divided by T, whereas in the usual formula for the sample covariance [see Equation (3.24)] the summation is divided by the number of observations in the summation, minus a degrees-of-freedom adjustment. The formula in Equation (14.5) is conventional for the purpose of computing the autocovariance. Equation (14.6) uses the assumption that var(Y_t) and var(Y_{t-j}) are the same—an implication of the assumption that Y is stationary, which is discussed in Section 14.4.

TABLE 14.2 First Four Sample Autocorrelations of the U.S. Inflation Rate and Its Change, 1960:I–2004:IV

	Autocorrelation of:	
Lag	Inflation Rate (Inf_t)	Change of Inflation Rate (ΔInf_t)
1	0.84	−0.26
2	0.76	−0.25
3	0.76	0.29
4	0.67	−0.06

Other Examples of Economic Time Series

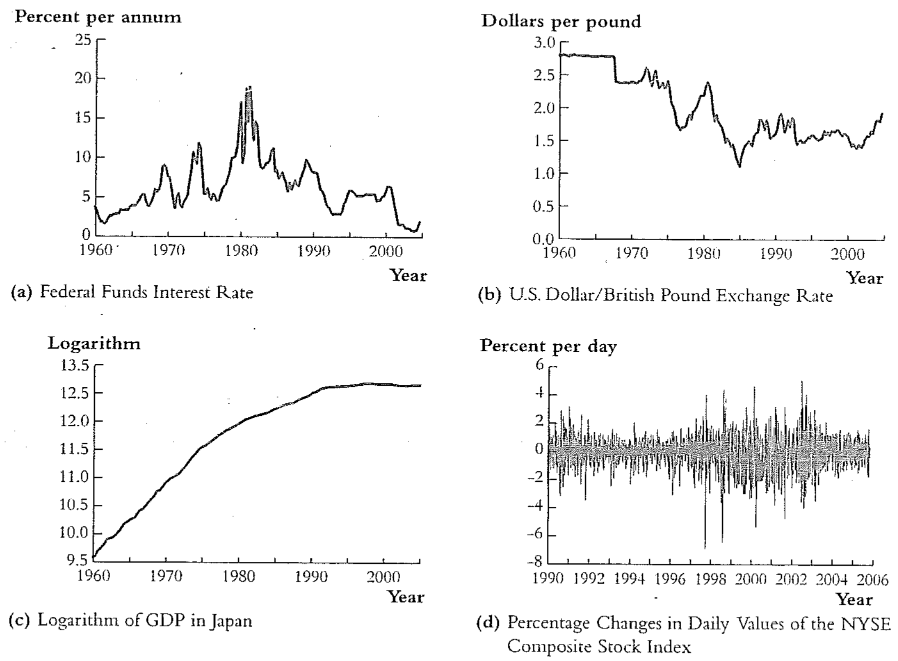
Economic time series differ greatly. Four examples of economic time series are plotted in Figure 14.2: the U.S. federal funds interest rate; the rate of exchange between the dollar and the British pound; the logarithm of Japanese gross domestic product; and the daily return on the Standard and Poor's 500 (S&P 500) stock market index.

The U.S. federal funds rate (Figure 14.2a) is the interest rate that banks pay to each other to borrow funds overnight. This rate is important because it is controlled by the Federal Reserve and is the Fed's primary monetary policy instrument. If you compare the plots of the federal funds rate and the rates of unemployment and inflation in Figure 14.1, you will see that sharp increases in the federal funds rate often have been associated with subsequent recessions.

The dollar/pound exchange rate (Figure 14.2b) is the price of a British pound (£) in U.S. dollars. Before 1972, the developed economies ran a system of fixed exchange rates—called the "Bretton Woods" system—under which governments worked to keep exchange rates from fluctuating. In 1972, inflationary pressures led to the breakdown of this system; thereafter, the major currencies were allowed to "float"; that is, their values were determined by the supply and demand for currencies in the market for foreign exchange. Prior to 1972, the exchange rate was approximately constant, with the exception of a single devaluation in 1968 in which the official value of the pound, relative to the dollar, was decreased to $2.40. Since 1972 the exchange rate has fluctuated over a very wide range.

Quarterly Japanese GDP (Figure 14.2c) is the total value of goods and services produced in Japan during a quarter. GDP is the broadest measure of total economic activity. The logarithm of the series is plotted in Figure 14.2c, and

FIGURE 14.2 Four Economic Time Series



(a) Federal Funds Interest Rate

(b) U.S. Dollar/British Pound Exchange Rate

(c) Logarithm of GDP in Japan

(d) Percentage Changes in Daily Values of the NYSE Composite Stock Index

The four time series have markedly different patterns. The federal funds rate (Figure 14.2a) has a pattern similar to price inflation. The exchange rate between the U.S. dollar and the British pound (Figure 14.2b) shows a discrete change after the 1972 collapse of the Bretton Woods system of fixed exchange rates. The logarithm of GDP in Japan (Figure 14.2c) shows relatively smooth growth, although the growth rate decreases in the 1970s and again in the 1990s. The daily percentage changes in the NYSE stock price index (Figure 14.2d) are essentially unpredictable, but its variance changes: This series shows "volatility clustering."

changes in this series can be interpreted as (fractional) growth rates. During the 1960s and early 1970s, Japanese GDP grew quickly, but this growth slowed in the late 1970s and 1980s. Growth slowed further during the 1990s, averaging only 1.2% per year from 1990 to 2004.

The NYSE Composite market index is a broad index of the share prices of all firms traded on the New York Stock Exchange. Figure 14.2d plots the daily percentage changes in this index for trading days from January 2, 1990, to November 11, 2005 (a total of 4003 observations). Unlike the other series in Figure 14.2, there is very little serial correlation in these daily percent changes: If there were, then you could predict them using past daily changes and make money by buying when you expect the market to rise and selling when you expect it to fall. Although the changes are essentially unpredictable, inspection of Figure 14.2d reveals patterns in their volatility. For example, the standard deviation of daily percentage changes was relatively large in 1990–1991 and 1998–2003, and relatively small in 1995 and 2005. This "volatility clustering" is found in many financial time series, and econometric models for modeling this special type of heteroskedasticity are taken up in Section 16.5.

14.3 Autoregressions

What will the rate of price inflation—the percentage increase in overall prices—be next year? Wall Street investors rely on forecasts of inflation when deciding how much to pay for bonds. Economists at central banks, like the U.S. Federal Reserve Bank, use inflation forecasts when they set monetary policy. Firms use inflation forecasts when they forecast sales of their products, and local governments use inflation forecasts when they develop their budgets for the upcoming year. In this section, we consider forecasts made using an **autoregression**, a regression model that relates a time series variable to its past values.

The First Order Autoregressive Model

If you want to predict the future of a time series, a good place to start is in the immediate past. For example, if you want to forecast the change in inflation from this quarter to the next, you might see whether inflation rose or fell last quarter. A systematic way to forecast the change in inflation, ΔInf_t, using the previous quarter's change, ΔInf_{t-1}, is to estimate an OLS regression of ΔInf_t on ΔInf_{t-1}. Estimated using data from 1962 to 2004, this regression is

$$\widehat{\Delta Inf_t} = \underset{(0.126)}{0.017} - \underset{(0.096)}{0.238}\Delta Inf_{t-1}, \tag{14.7}$$

where, as usual, standard errors are given in parentheses under the estimated coefficients, and $\widehat{\Delta Inf_t}$ is the predicted value of ΔInf_t based on the estimated regression line. The model in Equation (14.7) is called a first order autoregression: an autoregression because it is a regression of the series onto its own lag, ΔInf_{t-1}, and first order because only one lag is used as a regressor. The coefficient in Equation (14.7) is negative, so an increase in the inflation rate in one quarter is associated with a decline in the inflation rate in the next quarter.

A first order autoregression is abbreviated AR(1), where the "1" indicates that it is first order. The population AR(1) model for the series Y_t is

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t, \tag{14.8}$$

where u_t is an error term.

Forecasts and forecast errors. Suppose that you have historical data on Y and you want to forecast its future value. If Y_t follows the AR(1) model in Equation (14.8) and if β_0 and β_1 are known, then the forecast of Y_{T+1} based on Y_T is $\beta_0 + \beta_1 Y_T$.

In practice, β_0 and β_1 are unknown, so forecasts must be based on estimates of β_0 and β_1. We will use the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, which are constructed using historical data. In general, $\hat{Y}_{T+1|T}$ will denote the forecast of Y_{T+1} based on information through period T using a model estimated with data through period T. Accordingly, the forecast based on the AR(1) model in Equation (14.8) is

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T, \tag{14.9}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using historical data through time T.

The **forecast error** is the mistake made by the forecast; this is the difference between the value of Y_{T+1} that actually occurred and its forecasted value based on Y_T:

$$\text{Forecast error} = Y_{T+1} - \hat{Y}_{T+1|T}. \tag{14.10}$$

Forecasts versus predicted values. The forecast is *not* an OLS predicted value, and the forecast error is *not* an OLS residual. OLS predicted values are calculated for the observations in the sample used to estimate the regression. In contrast, the forecast is made for some date beyond the data set used to estimate the regression, so the data on the actual value of the forecasted dependent variable are not in the sample used to estimate the regression. Similarly, the OLS residual is the difference between the actual value of Y and its predicted value for observations in the sample, whereas the forecast error is the difference between the future value of Y, which is not contained in the estimation sample, and the forecast of that future value. Said differently, forecasts and forecast errors pertain to "out-of-sample" observations, whereas predicted values and residuals pertain to "in-sample" observations.

Root mean squared forecast error. The **root mean squared forecast error (RMSFE)** is a measure of the size of the forecast error, that is, of the magnitude of a typical mistake made using a forecasting model. The RMSFE is the square root of the mean squared forecast error:

$$\text{RMSFE} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}. \tag{14.11}$$

The RMSFE has two sources of error: the error arising because future values of u_t are unknown and the error in estimating the coefficients β_0 and β_1. If the first source of error is much larger than the second, as it can be if the sample size is large, then the RMSFE is approximately $\sqrt{\text{var}(u_t)}$, the standard deviation of the error u_t in the population autoregression [Equation (14.8)]. The standard deviation of u_t is in turn estimated by the standard error of the regression (SER; see Section 4.3). Thus, if uncertainty arising from estimating the regression coefficients is small enough to be ignored, the RMSFE can be estimated by the standard error of the regression. Estimation of the RMSFE including both sources of forecast error is taken up in Section 14.4.

Application to inflation. What is the forecast of inflation in the first quarter of 2005 (2005:I) that a forecaster would have made in 2004:IV, based on the estimated AR(1) model in Equation (14.7) (which was estimated using data through 2004:IV)? From Table 14.1, the inflation rate in 2004:IV was 3.5% (so $Inf_{2004:\text{IV}} =$ 3.5%), an increase of 1.9 percentage points from 2004:III (so $\Delta Inf_{2004:\text{IV}} = 1.9$). Plugging these values into Equation (14.7), the forecast of the change in inflation from 2004:IV to 2005:I is $\widehat{\Delta Inf}_{2005:\text{I}|2004:\text{IV}} = 0.017 - 0.238 \times \Delta Inf_{2004:\text{IV}} = 0.017 - 0.238 \times 1.9 \doteq -0.43 \cong -0.4$ (rounded to the nearest tenth). The predicted rate of inflation is the past rate of inflation plus its predicted change:

$$\widehat{Inf}_{T+1|T} = Inf_T + \widehat{\Delta Inf}_{T+1|T}. \tag{14.12}$$

Because $Inf_{2004:\text{IV}} = 3.5\%$ and the predicted change in the inflation rate from 2004:IV to 2005:I is −0.4, the predicted rate of inflation in 2005:I is $\widehat{Inf}_{2005:\text{I}|2004:\text{IV}} = Inf_{2004:\text{IV}} + \widehat{\Delta Inf}_{2005:\text{I}|2004:\text{IV}} = 3.5\% - 0.4\% = 3.1\%$. Thus, the AR(1) model forecasts that inflation will drop slightly from 3.5% in 2004:IV to 3.1% in 2005:I.

How accurate was this AR(1) forecast? From Table 14.1, the actual value of inflation in 2005:I was 2.4%, so the AR(1) forecast is high by 0.7 percentage point; that is, the forecast error is −0.7. The $\bar{R}^2$ of the AR(1) model in Equation (14.7) is only 0.05, so the lagged change of inflation explains a very small fraction of the variation in inflation in the sample used to fit the autoregression. This low $\bar{R}^2$ is consistent with the poor forecast of inflation in 2005:I produced using Equation (14.7). More generally, the low $\bar{R}^2$ suggests that this AR(1) model will forecast only a small amount of the variation in the change of inflation.

The standard error of the regression in Equation (14.7) is 1.65; ignoring uncertainty arising from estimation of the coefficients, our estimate of the RMSFE for forecasts based on Equation (14.7) therefore is 1.65 percentage points.

The p^{th}-Order Autoregressive Model

The AR(1) model uses Y_{t-1} to forecast Y_t, but doing so ignores potentially useful information in the more distant past. One way to incorporate this information is to include additional lags in the AR(1) model; this yields the p^{th}-order autoregressive, or AR(p), model.

The **p^{th}-order autoregressive model** [the **AR(p)** model] represents Y_t as a linear function of p of its lagged values; that is, in the AR(p) model, the regressors are $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$, plus an intercept. The number of lags, p, included in an AR(p) model is called the order, or lag length, of the autoregression.

For example, an AR(4) model of the change in inflation uses four lags of the change in inflation as regressors. Estimated by OLS over the period 1962–2004, the AR(4) model is

$$\begin{aligned}\widehat{\Delta Inf_t} = \underset{(0.12)}{0.02} - \underset{(0.09)}{0.26}\Delta Inf_{t-1} - \underset{(0.08)}{0.32}\Delta Inf_{t-2} + \underset{(0.08)}{0.16}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4}.\end{aligned} \tag{14.13}$$

The coefficients on the final three additional lags in Equation (14.13) are jointly significantly different from zero at the 5% significance level: The F-statistic is 6.91 (p-value < 0.001). This is reflected in an improvement in the $\bar{R}^2$ from 0.05 for the AR(1) model in Equation (14.7) to 0.18 for the AR(4) model. Similarly, the SER of the AR(4) model in Equation (14.13) is 1.52, an improvement over the SER of the AR(1) model, which is 1.65.

The AR(p) model is summarized in Key Concept 14.3.

Properties of the forecast and error term in the AR(p) model. The assumption that the conditional expectation of u_t is zero given past values of Y_t [that is, $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$] has two important implications.

The first implication is that the best forecast of Y_{T+1} based on its entire history depends on only the most recent p past values. Specifically, let $Y_{T+1|T} =$

KEY CONCEPT 14.3

Autoregressions

The p^{th}-order autoregressive model [the AR(p) model] represents Y_t as a linear function of p of its lagged values:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t, \quad (14.14)$$

where $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$. The number of lags p is called the order, or the lag length, of the autoregression.

$E(Y_{T+1}|Y_T, Y_{T-1}, \ldots)$ denote the conditional mean of Y_{T+1} given its entire history. Then $Y_{T+1|T}$ has the smallest RMSFE of any forecast based on the history of Y (Exercise 14.5). If Y_t follows an AR(p), then the best forcast of Y_{T+1} based on Y_T, $Y_{T-1}, \ldots$ is

$$Y_{T+1|T} = \beta_0 + \beta_1 Y_T + \beta_2 Y_{T-1} + \cdots + \beta_p Y_{T-p+1}, \quad (14.15)$$

which follows from the AR(p) model in Equation (14.14) and the assumption that $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$. In practice, the coefficients $\beta_0, \beta_1, \ldots, \beta_p$ are unknown, so actual forecasts from an AR(p) use Equation (14.15) with estimated coefficients.

The second implication is that the errors u_t are serially uncorrelated, a result that follows from Equation (2.27) (Exercise 14.5).

Application to inflation. What is the forecast of inflation in 2005:I using data through 2004:IV, based on the AR(4) model of inflation in Equation (14.13)? To compute this forecast, substitute the values of the change of inflation in each of the four quarters of 2004 into Equation (14.13): $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0.02 - 0.26\Delta Inf_{2004:IV} - 0.32\Delta Inf_{2004:III} + 0.16\Delta Inf_{2004:II} - 0.03\Delta Inf_{2004:I} = 0.02 - 0.26 \times 1.9 - 0.32 \times (-2.8) + 0.16 \times 0.6 - 0.03 \times 2.9 \cong 0.4$, where the 2004 values for the change of inflation are taken from the final column of Table 14.1.

The corresponding forecast of inflation in 2005:I is the value of inflation in 2004:IV, plus the forecasted change; that is, 3.5% + 0.4% = 3.9%. The forecast error is the actual value, 2.4%, minus the forecast, or 2.4% − 3.9% = −1.5, greater in absolute value than the AR(1) forecast error of −0.7 percentage point.

Can You Beat the Market? Part I

Have you ever dreamed of getting rich quick by beating the stock market? If you think that the market will be going up, you should buy stocks today and sell them later, before the market turns down. If you are good at forecasting swings in stock prices, then this active trading strategy will produce better returns than a passive "buy and hold" strategy in which you purchase stocks and just hang onto them. The trick, of course, is having a reliable forecast of future stock returns.

Forecasts based on past values of stock returns are sometimes called "momentum" forecasts: If the value of a stock rose this month, perhaps it has momentum and will also rise next month. If so, then returns will be autocorrelated and the autoregressive model will provide useful forecasts. You can implement a momentum-based strategy for a specific stock or for a stock index that measures the overall value of the market.

continued

TABLE 14.3 Autoregressive Models of Monthly Excess Stock Returns, 1960:1–2002:12

Dependent variable: excess returns on the CRSP value-weighted index.

	(1)	(2)	(3)
Specification	AR(1)	AR(2)	AR(4)
Regressors			
excess return$_{t-1}$	0.050 (0.051)	0.053 (0.051)	0.054 (0.051)
excess return$_{t-2}$		−0.053 (0.048)	−0.054 (0.048)
excess return$_{t-3}$			0.009 (0.050)
excess return$_{t-4}$			−0.016 (0.047)
Intercept	0.312 (0.197)	0.328 (0.199)	0.331 (0.202)
F-statistic for coefficients on lags of *excess return* (p-value)	0.968 (0.325)	1.342 (0.261)	0.707 (0.587)
$\bar{R}^2$	0.0006	0.0014	−0.0022

Notes: Excess returns are measured in percent per month. The data are described in Appendix 14.1. All regressions are estimated over 1960:1–2002:12 (T = 516 observations), with earlier observations used for initial values of lagged variables. Entries in the regressor rows are coefficients, with standard errors in parentheses. The final two rows report the F-statistic testing the hypothesis that the coefficients on lags of *excess return* in the regression are zero, with its p-value in parentheses, and the adjusted $\bar{R}^2$.

Table 14.3 presents autoregressive models of the excess return on a broad-based index of stock prices called the CRSP value-weighted index, using monthly data from 1960:1 to 2002:12. The monthly excess return is what you earn, in percentage terms, by purchasing a stock at the end of the previous month and selling it at the end of this month, minus what you would have earned had you purchased a safe asset (a U.S. Treasury bill). The return on the stock includes the capital gain (or loss) from the change in price, plus any dividends you receive during the month. The data are described further in Appendix 14.1.

Sadly, the results in Table 14.3 are negative. The coefficient on lagged returns in the AR(1) model is not statistically significant, and we cannot reject the null hypothesis that the coefficients on lagged returns are all zero in the AR(2) or AR(4) model. In fact, the adjusted R^2 of one of the models is negative and the other two are only slightly positive, suggesting that none of these models is useful for forecasting.

These negative results are consistent with the theory of efficient capital markets, which holds that excess returns should be unpredictable because stock prices already embody all currently available information. The reasoning is simple: If market participants think that a stock will have a positive excess return next month, then they will buy that stock now; but doing so will drive up the price of the stock to exactly the point at which there is no expected excess return. As a result, we should not be able to forecast future excess returns by using past publicly available information, nor can we, at least using the regressions in Table 14.3.

14.4 Time Series Regression with Additional Predictors and the Autoregressive Distributed Lag Model

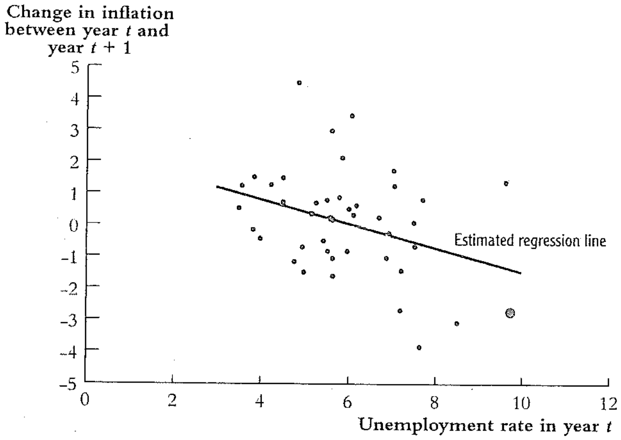
Economic theory often suggests other variables that could help to forecast the variable of interest. These other variables, or predictors, can be added to an autoregression to produce a time series regression model with multiple predictors. When other variables and their lags are added to an autoregression, the result is an autoregressive distributed lag model.

Forecasting Changes in the Inflation Rate Using Past Unemployment Rates

A high value of the unemployment rate tends to be associated with a future decline in the rate of inflation. This negative relationship, known as the short-run Phillips curve, is evident in the scatterplot of Figure 14.3, in which year-to-year changes in the rate of price inflation are plotted against the rate of unemployment in the previous year. For example, in 1982 the unemployment rate averaged 9.7%, and during the next year the rate of inflation fell by 2.9%. Overall, the correlation in Figure 14.3 is −0.36.

FIGURE 14.3 Scatterplot of Change in Inflation Between Year t and Year $t + 1$ versus the Unemployment Rate in Year t, 1961–2004

In 1982, the U.S. unemployment rate was 9.7% and the rate of inflation in 1983 fell by 2.9% (the large dot). In general, high values of the unemployment rate in year t tend to be followed by decreases in the rate of price inflation in the next year, year $t + 1$, with a correlation of −0.36.



The scatterplot in Figure 14.3 suggests that past values of the unemployment rate might contain information about the future course of inflation that is not already contained in past changes of inflation. This conjecture is readily checked by augmenting the AR(4) model in Equation (14.13) to include the first lag of the unemployment rate:

$$\begin{aligned}\widehat{\Delta Inf_t} = {} & \underset{(0.53)}{1.28} - \underset{(0.09)}{0.31}\Delta Inf_{t-1} - \underset{(0.09)}{0.39}\Delta Inf_{t-2} + \underset{(0.08)}{0.09}\Delta Inf_{t-3} \\ & - \underset{(0.09)}{0.08}\Delta Inf_{t-4} - \underset{(0.09)}{0.21}Unemp_{t-1}.\end{aligned} \tag{14.16}$$

The t-statistic on $Unemp_{t-1}$ is −2.23, so this term is significant at the 5% level. The $\bar{R}^2$ of this regression is 0.21, an improvement over the AR(4) $\bar{R}^2$ of 0.18.

The forecast of the change of inflation in 2005:I is obtained by substituting the 2004 values of the change of inflation into Equation (14.16), along with the value of the unemployment rate in 2004:IV (which is 5.4%); the resulting forecast is $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0.4$. Thus the forecast of inflation in 2005:I is 3.5% + 0.4% = 3.9%, and the forecast error is −1.5%.

If one lag of the unemployment rate is helpful for forecasting inflation, several lags might be even more helpful; adding three more lags of the unemployment rate yields

$$\begin{aligned}\widehat{\Delta Inf_t} = {} & \underset{(0.44)}{1.30} - \underset{(0.08)}{0.42}\Delta Inf_{t-1} - \underset{(0.09)}{0.37}\Delta Inf_{t-2} + \underset{(0.08)}{0.06}\Delta Inf_{t-3} - \underset{(0.08)}{0.04}\Delta Inf_{t-4} \\ & - \underset{(0.46)}{2.64}Unemp_{t-1} + \underset{(0.86)}{3.04}Unemp_{t-2} - \underset{(0.89)}{0.38}Unemp_{t-3} - \underset{(0.45)}{0.25}Unemp_{t-4}.\end{aligned} \tag{14.17}$$

The F-statistic testing the joint significance of the second through fourth lags of the unemployment rate is 10.76 (p-value < 0.001), so they are jointly significant. The $\bar{R}^2$ of the regression in Equation (14.17) is 0.34, a solid improvement over 0.21 for Equation (14.16). The F-statistic on all the unemployment coefficients is 8.91 (p-value < 0.001), indicating that this model represents a statistically significant improvement over the AR(4) model of Section 14.3 [Equation (14.13)]. The standard error of the regression in Equation (14.17) is 1.36, a substantial improvement over the SER of 1.52 for the AR(4).

The forecasted change in inflation from 2004:IV to 2005:I using Equation (14.17) is computed by substituting the values of the variables into the equation. The unemployment rate was 5.7% in 2004:I, 5.6% in 2004:II, and 5.4% in 2004:III and 2004:IV. The forecast of the change in inflation from 2004:IV to 2005:I, based on Equation (14.17), is

$$\begin{aligned}\widehat{\Delta Inf}_{2005:I|2004:IV} = {} & 1.30 - 0.42 \times 1.9 - 0.37 \times (-2.8) + 0.06 \times 0.6 - 0.04 \\ & \times 2.9 - 2.66 \times 5.4 + 0.34 \times 5.4 - 0.38 \times 5.6 - 0.25 \times 5.7 = 0.1.\end{aligned} \tag{14.18}$$

Thus the forecast of inflation in 2005:I is 3.5% + 0.1% = 3.6%. The forecast error is −1.2.

The autoregressive distributed lag model. Each model in Equations (14.16) and (14.17) is an **autoregressive distributed lag (ADL) model**: "autoregressive" because lagged values of the dependent variable are included as regressors, as in an autoregression, and "distributed lag" because the regression also includes multiple lags (a "distributed lag") of an additional predictor. In general, an autoregressive distributed lag model with p lags of the dependent variable Y_t and q lags of an additional predictor X_t is called an **ADL(p, q)** model. In this notation, the model in Equation (14.16) is an ADL(4,1) model and the model in Equation (14.17) is an ADL(4,4) model.

The Autoregressive Distributed Lag Model

KEY CONCEPT 14.4

The autoregressive distributed lag model with p lags of Y_t and q lags of X_t, denoted ADL(p, q), is

$$\begin{aligned}Y_t = {} & \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\ & + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t,\end{aligned} \tag{14.19}$$

where $\beta_0, \beta_1, \ldots, \beta_p, \delta_1, \ldots, \delta_q$ are unknown coefficients and u_t is the error term with $E(u_t|Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots) = 0$.

The autoregressive distributed lag model is summarized in Key Concept 14.4. With all these regressors, the notation in Equation (14.19) is somewhat cumbersome, and alternative optional notation, based on the so-called lag operator, is presented in Appendix 14.3.

The assumption that the errors in the ADL model have a conditional mean of zero given all past values of Y and X, that is, that $E(u_t|Y_{t-1}, Y_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots) = 0$, implies that no additional lags of either Y or X belong in the ADL model. In other words, the lag lengths p and q are the true lag lengths, and the coefficients on additional lags are zero.

The ADL model contains lags of the dependent variable (the autoregressive component) and a distributed lag of a single additional predictor, X. In general, however, forecasts can be improved by using multiple predictors. But before turning to the general time series regression model with multiple predictors, we first introduce the concept of stationarity, which will be used in that discussion.

Stationarity

Regression analysis of time series data necessarily uses data from the past to quantify historical relationships. If the future is like the past, then these historical relationships can be used to forecast the future. But if the future differs fundamentally from the past, then those historical relationships might not be reliable guides to the future.

In the context of time series regression, the idea that historical relationships can be generalized to the future is formalized by the concept of **stationarity**. The

KEY CONCEPT 14.5

Stationarity

A time series Y_t is *stationary* if its probability distribution does not change over time, that is, if the joint distribution of $(Y_{s+1}, Y_{s+2}, \ldots, Y_{s+T})$ does not depend on s regardless of the value of T; otherwise, Y_t is said to be *nonstationary*. A pair of time series, X_t and Y_t, are said to be *jointly stationary* if the joint distribution of $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \ldots, X_{s+T}, Y_{s+T})$ does not depend on s regardless of the value of T. Stationarity requires the future to be like the past, at least in a probabilistic sense.

precise definition of stationarity, given in Key Concept 14.5, is that the probability distribution of the time series variable does not change over time.

Time Series Regression with Multiple Predictors

The general time series regression model with multiple predictors extends the ADL model to include multiple predictors and their lags. The model is summarized in Key Concept 14.6. The presence of multiple predictors and their lags leads to double subscripting of the regression coefficients and regressors.

The time series regression model assumptions. The assumptions in Key Concept 14.6 modify the four least squares assumptions of the multiple regression model for cross-sectional data (Key Concept 6.4) for time series data.

The first assumption is that u_t has conditional mean zero, given all the regressors *and* the additional lags of the regressors beyond the lags included in the regression. This assumption extends the assumption used in the AR and ADL models and implies that the best forecast of Y_t using all past values of Y and the X's is given by the regression in Equation (14.20).

The second least squares assumption for cross-sectional data (Key Concept 6.4) is that $(X_{1i}, \ldots, X_{ki}, Y_i), i = 1, \ldots, n$, are independently and identically distributed (i.i.d.). The second assumption for time series regression replaces the i.i.d. assumption by a more appropriate one with two parts. Part (a) is that the data are drawn from a stationary distribution so that the distribution of the data today is the same as its distribution in the past. This assumption is a time series version of the "identically distributed" part of the i.i.d. assumption: The cross-sectional requirement of each draw being identically distributed is replaced by the time series requirement that the joint distribution of the variables, *including lags*, does not change over time. In practice, many economic time series appear to be nonstationary, which means that this assumption can fail to hold in applications. If the time series variables are nonstationary, then one or more problems can arise in time series regression: The forecast can be biased, the forecast can be inefficient (there can be alternative forecasts based on the same data with lower variance), or conventional OLS-based statistical inferences (for example, performing a hypothesis test by comparing the OLS t-statistic to ±1.96) can be misleading. Precisely which of these problems occurs, and its remedy, depends on the source of the nonstationarity. In Sections 14.6 and 14.7, we study the problems posed by, tests for, and solutions to two empirically important types of nonstationarity in economic time series, trends and breaks. For now, however, we simply assume that the series are jointly stationary and accordingly focus on regression with stationary variables.

KEY CONCEPT 14.6

Time Series Regression with Multiple Predictors

The general time series regression model allows for k additional predictors, where q_1 lags of the first predictor are included, q_2 lags of the second predictor are included, and so forth:

$$\begin{aligned} Y_t = \beta_0 &+ \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \\ &+ \delta_{11} X_{1t-1} + \delta_{12} X_{1t-2} + \cdots + \delta_{1q_1} X_{1t-q_1} \\ &+ \cdots + \delta_{k1} X_{kt-1} + \delta_{k2} X_{kt-2} + \cdots + \delta_{kq_k} X_{kt-q_k} + u_t, \end{aligned} \qquad (14.20)$$

where

1. $E(u_t | Y_{t-1}, Y_{t-2}, \ldots, X_{1t-1}, X_{1t-2}, \ldots, X_{kt-1}, X_{kt-2}, \ldots) = 0$;
2. (a) The random variables $(Y_t, X_{1t}, \ldots, X_{kt})$ have a stationary distribution, and (b) $(Y_t, X_{1t}, \ldots, X_{kt})$ and $(Y_{t-j}, X_{1t-j}, \ldots, X_{kt-j})$ become independent as j gets large;
3. Large outliers are unlikely: $X_{1t}, \ldots, X_{kt}$ and Y_t have nonzero, finite fourth moments; and
4. There is no perfect multicollinearity.

Part (b) of the second assumption requires that the random variables become independently distributed when the amount of time separating them becomes large. This replaces the cross-sectional requirement that the variables be independently distributed from one observation to the next with the time series requirement that they be independently distributed when they are separated by long periods of time. This assumption is sometimes referred to as **weak dependence**, and it ensures that in large samples there is sufficient randomness in the data for the law of large numbers and the central limit theorem to hold. We do not

KEY CONCEPT 14.7

Granger Causality Tests (Tests of Predictive Content)

The Granger causality statistic is the F-statistic testing the hypothesis that the coefficients on all the values of one of the variables in Equation (14.20) (for example, the coefficients on $X_{1t-1}, X_{1t-2}, \ldots, X_{1t-q_1}$) are zero. This null hypothesis implies that these regressors have no predictive content for Y_t beyond that contained in the other regressors, and the test of this null hypothesis is called the Granger causality test.

provide a precise mathematical statement of the weak dependence condition; rather, the reader is referred to Hayashi (2000, Chapter 2).

The third assumption, which is the same as the third least squares assumption for cross-sectional data, is that large outliers are unlikely, made mathematically precise by the assumption that all the variables have nonzero finite fourth moments.

Finally, the fourth assumption, which is also the same as for cross-sectional data, is that the regressors are not perfectly multicollinear.

Statistical inference and the Granger causality test. Under the assumptions of Key Concept 14.6, inference on the regression coefficients using OLS proceeds in the same way as it usually does using cross-sectional data.

One useful application of the F-statistic in time series forecasting is to test whether the lags of one of the included regressors has useful predictive content, above and beyond the other regressors in the model. The claim that a variable has no predictive content corresponds to the null hypothesis that the coefficients on all lags of that variable are zero. The F-statistic testing this null hypothesis is called the **Granger causality statistic**, and the associated test is called a **Granger causality test** (Granger, 1969). This test is summarized in Key Concept 14.7.

Granger causality has little to do with causality in the sense that it is used elsewhere in this book. In Chapter 1, causality was defined in terms of an ideal randomized controlled experiment, in which different values of X are applied experimentally and we observe the subsequent effect on Y. In contrast, Granger causality means that if X Granger-causes Y, then X is a useful predictor of Y, given the other variables in the regression. While "Granger predictability" is a more accurate term than "Granger causality," the latter has become part of the jargon of econometrics.

As an example, consider the relationship between the change in the inflation rate and its past values and past values of the unemployment rate. Based on the OLS estimates in Equation (14.17), the F-statistic testing the null hypothesis that the coefficients on all four lags of the unemployment rate are zero is 8.91 (p-value < 0.001): In the jargon of Key Concept 14.7, we can conclude (at the 1% significance level) that the unemployment rate Granger-causes changes in the inflation rate. This *does not* necessarily mean that a change in the unemployment rate will cause—in the sense of Chapter 1—a subsequent change in the inflation rate. It *does* mean that the past values of the unemployment rate appear to contain information that is useful for forecasting changes in the inflation rate, beyond that contained in past values of the inflation rate.

Forecast Uncertainty and Forecast Intervals

In any estimation problem, it is good practice to report a measure of the uncertainty of that estimate, and forecasting is no exception. One measure of the uncertainty of a forecast is its root mean square forecast error. Under the additional assumption that the errors u_t are normally distributed, the RMSFE can be used to construct a forecast interval, that is, an interval that contains the future value of the variable with a certain probability.

Forecast uncertainty. The forecast error consists of two components: uncertainty arising from estimation of the regression coefficients and uncertainty associated with the future unknown value of u_t. For regression with few coefficients and many observations, the uncertainty arising from future u_t can be much larger than the uncertainty associated with estimation of the parameters. In general, however, both sources of uncertainty are important, so we now develop an expression for the RMSFE that incorporates these two sources of uncertainty.

To keep the notation simple, consider forecasts of Y_{T+1} based on an ADL(1,1) model with a single predictor, that is, $Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + u_t$, and suppose that u_t is homoskedastic. The forecast is $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\delta}_1 X_T$, and the forecast error is

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T]. \qquad (14.21)$$

Because u_{T+1} has conditional mean zero and is homoskedastic, u_{T+1} has variance σ_u^2 and is uncorrelated with the final expression in brackets in Equation (14.21). Thus the mean squared forecast error (MSFE) is

$$\begin{aligned}\text{MSFE} &= E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] \\ &= \sigma_u^2 + \text{var}[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T], \qquad (14.22)\end{aligned}$$

and the RMSFE is the square root of the MSFE.

Estimation of the MSFE entails estimation of the two parts in Equation (14.22). The first term, σ_u^2, can be estimated by the square of the standard error of the regression, as discussed in Section 14.3. The second term requires estimating the variance of a weighted average of the regression coefficients, and methods for doing so were discussed in Section 8.1 [see the discussion following Equation (8.7)].

An alternative method for estimating the MSFE is to use the variance of pseudo out-of-sample forecasts, a procedure discussed in Section 14.7.

Forecast intervals. A forecast interval is like a confidence interval except that it pertains to a forecast. That is, a 95% **forecast interval** is an interval that contains the future value of the series in 95% of repeated applications.

One important difference between a forecast interval and a confidence interval is that the usual formula for a 95% confidence interval (the estimator ± 1.96 standard errors) is justified by the central limit theorem and therefore holds for a wide range of distributions of the error term. In contrast, because the forecast error in Equation (14.21) includes the future value of the error u_{T+1}, to compute a forecast interval requires either estimating the distribution of the error term or making some assumption about that distribution.

In practice, it is convenient to assume that u_{T+1} is normally distributed. If so, Equation (14.21) and the central limit theorem applied to $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\delta}_1$ imply that the forecast error is the sum of two independent, normally distributed terms, so the forecast error is itself normally distributed with variance equaling the MSFE. It follows that a 95% confidence interval is given by $\hat{Y}_{T+1|T} \pm 1.96SE(Y_{T+1} - \hat{Y}_{T+1|T})$, where $SE(Y_{T+1} - \hat{Y}_{T+1|T})$ is an estimator of the RMSFE.

This discussion has focused on the case that the error term, u_{T+1}, is homoskedastic. If instead u_{T+1} is heteroskedastic, then one needs to develop a model of the heteroskedasticity so that the term σ_u^2 in Equation (14.22) can be estimated, given the most recent values of Y and X, and methods for modeling this conditional heteroskedasticity are presented in Section 16.5.

Because of uncertainty about future events—that is, uncertainty about u_{T+1}—95% forecast intervals can be so wide that they have limited use in decision making. Professional forecasters therefore often report forecast intervals that are tighter than 95%, for example, one standard error forecast intervals (which are 68% forecast intervals if the errors are normally distributed). Alternatively, some forecasters report multiple forecast intervals, as is done by the economists at the Bank of England when they publish their inflation forecasts (see "The River of Blood" on the following page).

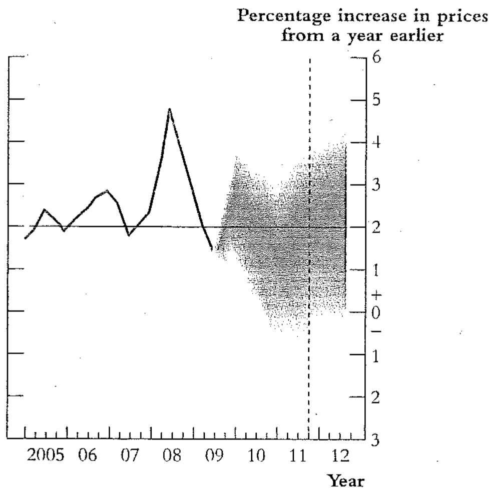
The River of Blood

As part of its efforts to inform the public about monetary policy decisions, the Bank of England regularly publishes forecasts of inflation. These forecasts combine output from econometric models maintained by professional econometricians at the bank with the expert judgment of the members of the bank's senior staff and Monetary Policy Committee. The forecasts are presented as a set of forecast intervals designed to reflect what these economists consider to be the range of probable paths that inflation might take. In its *Inflation Report*, the bank prints these ranges in red, with the darkest red reserved for the central band. Although the bank prosaically refers to this as the "fan chart," the press has called these spreading shades of red the "river of blood."

The river of blood for November 2009 is shown in Figure 14.4 (in this figure the blood is blue, not red, so you will need to use your imagination). This chart shows that, as of November 2009, the bank's economists expected the rate of inflation to increase sharply to roughly 3% in early 2010, fall to approximately 1% by the end of 2010, and then climb steadily back to 2% by 2012. The economists expressed considerable uncertainty about the forecast, however. They cited an increase in the VAT (sales tax) as an important factor increasing inflation in the short run and discussed uncertainty associated with inflation's response to the slack in economy and the timing and strength of the economic recovery as important sources of inflation uncertainty. As it turns out, their near-

FIGURE 14.4 The River of Blood

The Bank of England's fan chart for November 2009 shows forecast ranges for inflation. The dashed line indicates the second quarter of 2011, two years after the release of the report.



continued

term forecast was very close to actual inflation—inflation in the second quarter of 2010 was 3.5%.

The Bank of England has been a pioneer in the movement toward greater openness by central banks, and other central banks now also publish inflation forecasts. The decisions made by monetary policymakers are difficult ones and affect the lives—and wallets—of many of their fellow citizens. In a democracy in the information age, reasoned the economists at the Bank of England, it is particularly important for citizens to understand the bank's economic outlook and the reasoning behind its difficult decisions.

To see the river of blood in its original red hue, visit the Bank of England's Web site at **www.bankofengland.co.uk.** To learn more about the performance of the Bank of England inflation forecasts, see Clements (2004).

14.5 Lag Length Selection Using Information Criteria

The estimated inflation regressions in Sections 14.3 and 14.4 have either one or four lags of the predictors. One lag makes some sense, but why four? More generally, how many lags should be included in a time series regression? This section discusses statistical methods for choosing the number of lags, first in an autoregression and then in a time series regression model with multiple predictors.

Determining the Order of an Autoregression

In practice, choosing the order p of an autoregression requires balancing the marginal benefit of including more lags against the marginal cost of additional estimation uncertainty. On the one hand, if the order of an estimated autoregression is too low, you will omit potentially valuable information contained in the more distant lagged values. On the other hand, if it is too high, you will be estimating more coefficients than necessary, which in turn introduces additional estimation error into your forecasts.

The F-statistic approach. One approach to choosing p is to start with a model with many lags and to perform hypothesis tests on the final lag. For example, you might start by estimating an AR(6) and test whether the coefficient on the sixth lag is significant at the 5% level; if not, drop it and estimate an AR(5), test the coefficient on the fifth lag, and so forth. The drawback to this method is that it will produce too large a model, at least some of the time: Even if the true AR order is five, so the sixth coefficient is zero, a 5% test using the t-statistic will incorrectly reject this null hypothesis 5% of the time just by chance. Thus, when the true value of p is five, this method will estimate p to be six 5% of the time.

The BIC. A way around this problem is to estimate p by minimizing an "information criterion." One such information criterion is the **Bayes information criterion (BIC)**, also called the *Schwarz information criterion (SIC)*, which is

$$\text{BIC}(p) = \ln\left[\frac{SSR(p)}{T}\right] + (p+1)\frac{\ln(T)}{T}, \qquad (14.23)$$

where $SSR(p)$ is the sum of squared residuals of the estimated AR(p). The BIC estimator of p, $\hat{p}$, is the value that minimizes BIC(p) among the possible choices $p = 0, 1, \dots, p_{max}$, where p_{max} is the largest value of p considered and $p = 0$ corresponds to the model that contains only an intercept.

The formula for the BIC might look a bit mysterious at first, but it has an intuitive appeal. Consider the first term in Equation (14.23). Because the regression coefficients are estimated by OLS, the sum of squared residuals necessarily decreases (or at least does not increase) when you add a lag. In contrast, the second term is the number of estimated regression coefficients (the number of lags, p, plus one for the intercept) times the factor $\ln(T)/T$. This second term increases when you add a lag. The BIC trades off these two forces so that the number of lags that minimizes the BIC is a consistent estimator of the true lag length. The mathematics of this argument is given in Appendix 14.5.

As an example, consider estimating the AR order for an autoregression of the change in the inflation rate. The various steps in the calculation of the BIC are carried out in Table 14.4 for autoregressions of maximum order six ($p_{max} = 6$). For example, for the AR(1) model in Equation (14.7), $SSR(1)/T = 2.737$, so $\ln[SSR(1)/T] = 1.007$. Because $T = 172$ (43 years, four quarters per year), $\ln(T)/T = 0.030$ and $(p+1)\ln(T)/T = 2 \times 0.030 = 0.060$. Thus BIC(1) = 1.007 + 0.060 = 1.067.

The BIC is smallest when $p = 2$ in Table 14.4. Thus the BIC estimate of the lag length is 2. As can be seen in Table 14.4, as the number of lags increases the R^2 increases and the SSR decreases. The increase in the R^2 is large from one to two lags, smaller from two to three, and quite small from three to four. The BIC helps decide precisely how large the increase in the R^2 must be to justify including the additional lag.

The AIC. The BIC is not the only information criterion; another is the **Akaike information criterion (AIC):**

$$\text{AIC}(p) = \ln\left[\frac{SSR(p)}{T}\right] + (p+1)\frac{2}{T}. \qquad (14.24)$$

TABLE 14.4 The Bayes Information Criterion (BIC) and the R^2 for Autoregressive Models of U.S. Inflation, 1962–2004

p	$SSR(p)/T$	$\ln[SSR(p)/T]$	$(p+1)\ln(T)/T$	BIC(p)	R^2
0	2.900	1.065	0.030	1.095	0.000
1	2.737	1.007	0.060	1.067	0.056
2	2.375	0.865	0.090	0.955	0.181
3	2.311	0.838	0.120	0.957	0.203
4	2.309	0.837	0.150	0.986	0.204
5	2.308	0.836	0.180	1.016	0.204
6	2.308	0.836	0.209	1.046	0.204

The difference between the AIC and the BIC is that the term "$\ln(T)$" in the BIC is replaced by "2" in the AIC, so the second term in the AIC is smaller. For example, for the 172 observations used to estimate the inflation autoregressions, $\ln(T) = \ln(172) = 5.15$, so that the second term for the BIC is more than twice as large as the term in AIC. Thus a smaller decrease in the *SSR* is needed in the AIC to justify including another lag. As a matter of theory, the second term in the AIC is not large enough to ensure that the correct lag length is chosen, even in large samples, so the AIC estimator of p is not consistent. As is discussed in Appendix 14.5, in large samples the AIC will overestimate p with nonzero probability.

Despite this theoretical blemish, the AIC is widely used in practice. If you are concerned that the BIC might yield a model with too few lags, the AIC provides a reasonable alternative.

A note on calculating information criteria. How well two estimated regressions fit the data is best assessed when they are estimated using the same data sets. Because the BIC and AIC are formal methods for making this comparison, the autoregressions under consideration should be estimated using the same observations. For example, in Table 14.4 all the regressions were estimated using data from 1962:I to 2004:IV, for a total of 172 observations. Because the autoregressions involve lags of the change of inflation, this means that earlier values of the change of inflation (values before 1962:I) were used as regressors for the preliminary observations. Said differently, the regressions examined in Table 14.4 each include observations on $\Delta Inf_t, \Delta Inf_{t-1}, \ldots, \Delta Inf_{t-p}$ for $t = 1962{:}\text{I}, \ldots, 2004{:}\text{IV}$, corresponding to 172 observations on the dependent variable and regressors, so $T = 172$ in Equations (14.23) and (14.24).

Lag Length Selection in Time Series Regression with Multiple Predictors

The trade-off involved with lag length choice in the general time series regression model with multiple predictors [Equation (14.20)] is similar to that in an autoregression: Using too few lags can decrease forecast accuracy because valuable information is lost, but adding lags increases estimation uncertainty. The choice of lags must balance the benefit of using additional information against the cost of estimating the additional coefficients.

The F-statistic approach. As in the univariate autoregression, one way to determine the number of lags to include is to use F-statistics to test joint hypotheses that sets of coefficients equal zero. For example, in the discussion of Equation (14.17), we tested the hypothesis that the coefficients on the second through fourth lags of the unemployment rate equal zero against the alternative that they are nonzero; this hypothesis was rejected at the 1% significance level, lending support to the longer-lag specification. If the number of models being compared is small, then this F-statistic method is easy to use. In general, however, the F-statistic method can produce models that are too large, in the sense that the true lag order is overestimated.

Information criteria. As in an autoregression, the BIC and AIC can be used to estimate the number of lags and variables in the time series regression model with multiple predictors. If the regression model has K coefficients (including the intercept), the BIC is

$$\text{BIC}(K) = \ln\left[\frac{SSR(K)}{T}\right] + K\frac{\ln(T)}{T}. \tag{14.25}$$

The AIC is defined in the same way, but with 2 replacing $\ln(T)$ in Equation (14.25). For each candidate model, the BIC (or AIC) can be evaluated, and the model with the lowest value of the BIC (or AIC) is the preferred model, based on the information criterion.

There are two important practical considerations when using an information criterion to estimate the lag lengths. First, as is the case for the autoregression, all the candidate models must be estimated over the same sample; in the notation of Equation (14.25), the number of observations used to estimate the model, T, must be the same for all models. Second, when there are multiple predictors, this approach is computationally demanding because it requires computing many different models (many combinations of the lag parameters). In practice, a convenient shortcut is to require all the regressors to have the same number of lags, that is, to require that $p = q_1 = \cdots = q_k$, so that only $p_{max} + 1$ models need to be compared (corresponding to $p = 0, 1, \ldots, p_{max}$).

14.6 Nonstationarity I: Trends

In Key Concept 14.6, it was assumed that the dependent variable and the regressors are stationary. If this is not the case, that is, if the dependent variable and/or regressors are nonstationary, then conventional hypothesis tests, confidence intervals, and forecasts can be unreliable. The precise problem created by nonstationarity, and the solution to that problem, depends on the nature of that nonstationarity.

In this and the next section, we examine two of the most important types of nonstationarity in economic time series data: trends and breaks. In each section, we first describe the nature of the nonstationarity and then discuss the consequences for time series regression if this type of nonstationarity is present but is ignored. We next present tests for nonstationarity and discuss remedies for, or solutions to, the problems caused by that particular type of nonstationarity. We begin by discussing trends.

What Is a Trend?

A **trend** is a persistent long-term movement of a variable over time. A time series variable fluctuates around its trend.

Inspection of Figure 14.1a suggests that the U.S. inflation rate has a trend consisting of a general upward tendency through 1982 and a downward tendency thereafter. The series in Figures 14.2a, b, and c also have trends, but their trends are quite different. The trend in the U.S. federal funds interest rate is similar to the trend in the U.S. inflation rate. The \$/£ exchange rate clearly had a prolonged downward trend after the collapse of the fixed exchange rate system in 1972. The logarithm of Japanese GDP has a complicated trend: fast growth at first, then moderate growth, and finally slow growth.

Deterministic and stochastic trends. There are two types of trends seen in time series data: deterministic and stochastic. A **deterministic trend** is a nonrandom function of time. For example, a deterministic trend might be linear in time; if inflation had a deterministic linear trend so that it increased by 0.1 percentage point per quarter, this trend could be written as $0.1t$, where t is measured in quarters. In contrast, a **stochastic trend** is random and varies over time. For example, a stochastic trend in inflation might exhibit a prolonged period of increase followed by a prolonged period of decrease, like the inflation trend in Figure 14.1.

Like many econometricians, we think it is more appropriate to model economic time series as having stochastic rather than deterministic trends. Economics is complicated stuff. It is hard to reconcile the predictability implied by a deterministic trend with the complications and surprises faced year after year by workers, businesses, and governments. For example, although U.S. inflation rose through the 1970s, it was neither destined to rise forever nor destined to fall again. Rather, the slow rise of inflation is now understood to have occurred because of bad luck and monetary policy mistakes, and its taming was in large part a consequence of tough decisions made by the Board of Governors of the Federal Reserve. Similarly, the \$/£ exchange rate trended down from 1972 to 1985 and subsequently drifted up, but these movements too were the consequences of complex economic forces; because these forces change unpredictably, these trends are usefully thought of as having a large unpredictable, or random, component.

For these reasons, our treatment of trends in economic time series focuses on stochastic rather than deterministic trends, and when we refer to "trends" in time series data we mean stochastic trends unless we explicitly say otherwise. This section presents the simplest model of a stochastic trend, the random walk model; other models of trends are discussed in Section 16.3.

The random walk model of a trend. The simplest model of a variable with a stochastic trend is the random walk. A time series Y_t is said to follow a **random walk** if the change in Y_t is i.i.d., that is, if

$$Y_t = Y_{t-1} + u_t, \qquad (14.26)$$

where u_t is i.i.d. We will, however, use the term *random walk* more generally to refer to a time series that follows Equation (14.26), where u_t has conditional mean zero; that is, $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$.

The basic idea of a random walk is that the value of the series tomorrow is its value today, plus an unpredictable change: Because the path followed by Y_t consists of random "steps" u_t, that path is a "random walk." The conditional mean of Y_t based on data through time $t - 1$ is Y_{t-1}; that is, because $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$, $E(Y_t|Y_{t-1}, Y_{t-2}, \ldots) = Y_{t-1}$. In other words, if Y_t follows a random walk, then the best forecast of tomorrow's value is its value today.

Some series, such as the logarithm of Japanese GDP in Figure 14.2c, have an obvious upward tendency, in which case the best forecast of the series must include an adjustment for the tendency of the series to increase. This adjustment leads to an extension of the random walk model to include a tendency to move, or "drift," in one direction or the other. This extension is referred to as a **random walk with drift**:

$$Y_t = \beta_0 + Y_{t-1} + u_t, \tag{14.27}$$

where $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$ and β_0 is the "drift" in the random walk. If β_0 is positive, then Y_t increases on average. In the random walk with drift model, the best forecast of the series tomorrow is the value of the series today, plus the drift β_0.

The random walk model (with drift as appropriate) is simple yet versatile, and it is the primary model for trends used in this book.

A random walk is nonstationary. If Y_t follows a random walk, then it is not stationary: The variance of a random walk increases over time, so the distribution of Y_t changes over time. One way to see this is to recognize that, because u_t is uncorrelated with Y_{t-1} in Equation (14.26), $\text{var}(Y_t) = \text{var}(Y_{t-1}) + \text{var}(u_t)$; for Y_t to be stationary, $\text{var}(Y_t)$ cannot depend on time, so in particular $\text{var}(Y_t) = \text{var}(Y_{t-1})$ must hold, but this can happen only if $\text{var}(u_t) = 0$. Another way to see this is to imagine that Y_t starts out at zero; that is, $Y_0 = 0$. Then $Y_1 = u_1$, $Y_2 = u_1 + u_2$, and so forth so that $Y_t = u_1 + u_2 + \cdots + u_t$. Because u_t is serially uncorrelated, $\text{var}(Y_t) = \text{var}(u_1 + u_2 + \cdots + u_t) = t\sigma_u^2$. Thus the variance of Y_t depends on t; in fact, it increases as t increases. Because the variance of Y_t depends on t, its distribution depends on t; that is, it is nonstationary.

Because the variance of a random walk increases without bound, its *population* autocorrelations are not defined (the first autocovariance and variance are infinite, and the ratio of the two is not well defined). However, a feature of a random walk is that its *sample* autocorrelations tend to be very close to 1; in fact, the j^{th} sample autocorrelation of a random walk converges to 1 in probability.

Stochastic trends, autoregressive models, and a unit root. The random walk model is a special case of the AR(1) model [Equation (14.8)] in which $\beta_1 = 1$. In other words, if Y_t follows an AR(1) with $\beta_1 = 1$, then Y_t contains a stochastic trend and is nonstationary. If, however, $|\beta_1| < 1$ and u_t is stationary, then the joint distribution of Y_t and its lags does not depend on t (a result shown in Appendix 14.2), so Y_t is stationary.

The analogous condition for an AR(p) to be stationary is more complicated than the condition $|\beta_1| < 1$ for an AR(1). Its formal statement involves the roots of the polynomial, $1 - \beta_1 z - \beta_2 z^2 - \beta_3 z^3 - \cdots - \beta_p z^p$. (The roots of this polynomial are the values of z that satisfy $1 - \beta_1 z - \beta_2 z^2 - \beta_3 z^3 - \cdots - \beta_p z^p = 0$.) For an AR($p$) to be stationary, the roots of this polynomial must all be greater than 1 in absolute value. In the special case of an AR(1), the root is the value of z that solves $1 - \beta_1 z = 0$, so its root is $z = 1/\beta_1$. Thus the statement that the root be greater than 1 in absolute value is equivalent to $|\beta_1| < 1$.

If an AR(p) has a root that equals 1, the series is said to have a *unit autoregressive root* or, more simply, a **unit root**. If Y_t has a unit root, then it contains a stochastic trend. If Y_t is stationary (and thus does not have a unit root), it does not contain a stochastic trend. For this reason, we will use the terms *stochastic trend* and *unit root* interchangeably.

Problems Caused by Stochastic Trends

If a regressor has a stochastic trend (has a unit root), then the OLS estimator of its coefficient and its OLS t-statistic can have nonstandard (that is, nonnormal) distributions, even in large samples. We discuss three specific aspects of this problem: (1) The estimator of the autoregressive coefficient in an AR(1) is biased toward 0 if its true value is 1; (2) the t-statistic on a regressor with a stochastic trend can have a nonnormal distribution, even in large samples; and (3) an extreme example of the risks posed by stochastic trends is that two series that are independent will, with high probability, misleadingly appear to be related if they both have stochastic trends, a situation known as spurious regression.

Problem #1: Autoregressive coefficients that are biased toward zero. Suppose that Y_t follows the random walk in Equation (14.26) but this is unknown to the econometrician, who instead estimates the AR(1) model in Equation (14.8). Because Y_t is nonstationary, the least squares assumptions for time series regression in Key Concept 14.6 do not hold, so as a general matter we cannot rely on estimators and test statistics having their usual large-sample normal distributions. In fact, in this example the OLS estimator of the autoregressive coefficient, $\hat{\beta}_1$, is consistent, but it has a nonnormal distribution, even in large samples: The asymptotic distribution of $\hat{\beta}_1$ is shifted toward zero. The expected value of $\hat{\beta}_1$ is approximately $E(\hat{\beta}_1) = 1 - 5.3/T$. This results in a large bias in sample sizes typically encountered in economic applications. For example, 20 years of quarterly data contain 80 observations, in which case the expected value of $\hat{\beta}_1$ is $E(\hat{\beta}_1) = 1 - 5.3/80 = 0.934$. Moreover, this distribution has a long left tail: The 5% percentile of $\hat{\beta}_1$ is approximately $1 - 14.1/T$, which, for $T = 80$, corresponds to 0.824, so 5% of the time $\hat{\beta}_1 < 0.824$.

One implication of this bias toward zero is that if Y_t follows a random walk, then forecasts based on the AR(1) model can perform substantially worse than those based on the random walk model, which imposes the true value $\beta_1 = 1$. This conclusion also applies to higher-order autoregressions, in which there are forecasting gains from imposing a unit root (that is, from estimating the autoregression in first differences instead of in levels) when in fact the series contains a unit root.

Problem #2: Nonnormal distributions of t-statistics. If a regressor has a stochastic trend, then its usual OLS t-statistic can have a nonnormal distribution under the null hypothesis, even in large samples. This nonnormal distribution means that conventional confidence intervals are not valid and hypothesis tests cannot be conducted as usual. In general, the distribution of this t-statistic is not readily tabulated because the distribution depends on the relationship between the regressor in question and the other regressors. An important example of this problem arises in regressions that attempt to forecast stock returns using regressors that could have stochastic trends (see the box in Section 14.7, "Can You Beat the Market? Part II").

One important case in which it *is* possible to tabulate the distribution of the t-statistic when the regressor has a stochastic trend is in the context of an autoregression with a unit root. We return to this special case when we take up the problem of testing whether a time series contains a stochastic trend.

Problem #3: Spurious regression. Stochastic trends can lead two time series to appear related when they are not, a problem called **spurious regression**.

For example, U.S. inflation was steadily rising from the mid-1960s through the early 1980s, and at the same time Japanese GDP (plotted in logarithms in Figure 14.2c) was steadily rising. These two trends conspire to produce a regression that appears to be "significant" using conventional measures. Estimated by OLS using data from 1965 through 1981, this regression is

$$\widehat{U.S.\ Inflation_t} = \underset{(3.99)}{-37.78} + \underset{(0.36)}{3.83} \times \ln(Japanese\ GDP_t),\ \bar{R}^2 = 0.56. \quad (14.28)$$

The t-statistic on the slope coefficient exceeds 10, which by usual standards indicates a strong positive relationship between the two series, and the $\bar{R}^2$ is high. However, running this regression using data from 1982 through 2004 yields

$$\widehat{U.S.\ Inflation_t} = \underset{(10.41)}{31.20} - \underset{(0.80)}{2.17} \times \ln(Japanese\ GDP_t),\ \bar{R}^2 = 0.08. \quad (14.29)$$

The regressions in Equations (14.28) and (14.29) could hardly be more different. Interpreted literally, Equation (14.28) indicates a strong positive relationship, while Equation (14.29) indicates a weak, but apparently statistically significant, negative relationship.

The source of these conflicting results is that both series have stochastic trends. These trends happened to align from 1965 through 1981, but did not align from 1982 through 2004. There is, in fact, no compelling economic or political reason to think that the trends in these two series are related. In short, these regressions are spurious.

The regressions in Equations (14.28) and (14.29) illustrate empirically the theoretical point that OLS can be misleading when the series contain stochastic trends (see Exercise 14.6 for a computer simulation that demonstrates this result). One special case in which certain regression-based methods *are* reliable is when the trend component of the two series is the same, that is, when the series contain a *common* stochastic trend; if so, the series are said to be cointegrated. Econometric methods for detecting and analyzing cointegrated economic time series are discussed in Section 16.4.

Detecting Stochastic Trends: Testing for a Unit AR Root

Trends in time series data can be detected by informal and formal methods. The informal methods involve inspecting a time series plot of the data and computing the autocorrelation coefficients, as we did in Section 14.2. Because the first autocorrelation coefficient will be near 1 if the series has a stochastic trend, at least in large samples, a small first autocorrelation coefficient combined with a time series plot that has no apparent trend suggests that the series does not have a trend. If doubt remains, however, there are formal statistical procedures that can be used to test the hypothesis that there is a stochastic trend in the series against the alternative that there is no trend.

In this section, we use the Dickey–Fuller test (named after its inventors David Dickey and Wayne Fuller, 1979) to test for a stochastic trend. Although the Dickey–Fuller test is not the only test for a stochastic trend (another test is discussed in Section 16.3), it is the most commonly used test in practice and is one of the most reliable.

The Dickey–Fuller test in the AR(1) model. The starting point for the **Dickey–Fuller test** is the autoregressive model. As discussed earlier, the random walk in Equation (14.27) is a special case of the AR(1) model with $\beta_1 = 1$. If $\beta_1 = 1$, Y_t is nonstationary and contains a (stochastic) trend. Thus, within the AR(1) model, the hypothesis that Y_t has a trend can be tested by testing

$$H_0: \beta_1 = 1 \text{ vs. } H_1: \beta_1 < 1 \text{ in } Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t. \quad (14.30)$$

If $\beta_1 = 1$, the AR(1) has an autoregressive root of 1, so the null hypothesis in Equation (14.30) is that the AR(1) has a unit root, and the alternative is that it is stationary.

This test is most easily implemented by estimating a modified version of Equation (14.30) obtained by subtracting Y_{t-1} from both sides. Let $\delta = \beta_1 - 1$; then Equation (14.30) becomes

$$H_0\text{: } \delta = 0 \text{ vs. } H_1\text{: } \delta < 0 \text{ in } \Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t. \tag{14.31}$$

The OLS t-statistic testing $\delta = 0$ in Equation (14.31) is called the **Dickey–Fuller statistic.** The formulation in Equation (14.31) is convenient because regression software automatically prints out the t-statistic testing $\delta = 0$. Note that the Dickey–Fuller test is one-sided, because the relevant alternative is that Y_t is stationary, so $\beta_1 < 1$ or, equivalently, $\delta < 0$. The Dickey–Fuller statistic is computed using "nonrobust" standard errors, that is, the "homoskedasticity-only" standard errors presented in Appendix 5.1 [Equation (5.29) for the case of a single regressor and in Section 18.4 for the multiple regression model].[3]

The Dickey–Fuller test in the AR(p) model. The Dickey–Fuller statistic presented in the context of Equation (14.31) applies only to an AR(1). As discussed in Section 14.3, for some series the AR(1) model does not capture all the serial correlation in Y_t, in which case a higher-order autoregression is more appropriate.

The extension of the Dickey–Fuller test to the AR(p) model is summarized in Key Concept 14.8. Under the null hypothesis, $\delta = 0$ and ΔY_t is a stationary AR(p). Under the alternative hypothesis, $\delta < 0$ so that Y_t is stationary. Because the regression used to compute this version of the Dickey–Fuller statistic is augmented by lags of ΔY_t, the resulting t-statistic is referred to as the **augmented Dickey–Fuller (ADF) statistic.**

In general the lag length p is unknown, but it can be estimated using an information criterion applied to regressions of the form in Equation (14.32) for various values of p. Studies of the ADF statistic suggest that it is better to have too many lags than too few, so it is recommended to use the AIC instead of the BIC to estimate p for the ADF statistic.[4]

Testing against the alternative of stationarity around a linear deterministic time trend. The discussion so far has considered the null hypothesis that the series has a unit root and the alternative hypothesis that it is stationary. This alternative hypothesis of stationarity is appropriate for series, such as the rate of inflation, that do not exhibit long-term growth. But other economic time series, such as Japanese GDP (Figure 14.2c), exhibit long-run growth, and for such series the alternative of stationarity without a trend is inappropriate. Instead, a commonly used alternative is that the series are stationary around a deterministic time trend, that is, a trend that is a deterministic function of time.

One specific formulation of this alternative hypothesis is that the time trend is linear, that is, the trend is a linear function of t; thus the null hypothesis is that the series has a unit root, and the alternative is that it does not have a unit root but does have a deterministic time trend. The Dickey–Fuller regression must be modified to test the null hypothesis of a unit root against the alternative that it is stationary around a linear time trend. As summarized in Equation (14.33) in Key Concept 14.8, this is accomplished by adding a time trend (the regressor $X_t = t$) to the regression.

The Augmented Dickey–Fuller Test for a Unit Autoregressive Root

KEY CONCEPT

14.8

The augmented Dickey–Fuller (ADF) test for a unit autoregressive root tests the null hypothesis H_0: $\delta = 0$ against the one-sided alternative H_1: $\delta < 0$ in the regression

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t. \tag{14.32}$$

Under the null hypothesis, Y_t has a stochastic trend; under the alternative hypothesis, Y_t is stationary. The ADF statistic is the OLS t-statistic testing $\delta = 0$ in Equation (14.32).

If instead the alternative hypothesis is that Y_t is stationary around a deterministic linear time trend, then this trend, "t" (the observation number), must be added as an additional regressor, in which case the Dickey–Fuller regression becomes

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t, \tag{14.33}$$

where α is an unknown coefficient and the ADF statistic is the OLS t-statistic testing $\delta = 0$ in Equation (14.33).

The lag length p can be estimated using the BIC or AIC. When $p = 0$, lags of ΔY_t are not included as regressors in Equations (14.32) and (14.33), and the ADF test simplifies to the Dickey–Fuller test in the AR(1) model. The ADF statistic does *not* have a normal distribution, even in large samples. Critical values for the one-sided ADF test depend on whether the test is based on Equation (14.32) or (14.33) and are given in Table 14.5.

[3] Under the null hypothesis of a unit root, the usual "nonrobust" standard errors produce a t-statistic that is in fact robust to heteroskedasticity, a surprising and special result.

[4] See Stock (1994) and Haldrup and Jansson (2006) for reviews of simulation studies of the finite-sample properties of the Dickey–Fuller and other unit root test statistics.

A linear time trend is not the only way to specify a deterministic time trend; for example, the deterministic time trend could be quadratic, or it could be linear but have breaks (that is, be linear with slopes that differ in two parts of the sample). The use of alternatives like these with nonlinear deterministic trends should be motivated by economic theory. For a discussion of unit root tests against stationarity around nonlinear deterministic trends, see Maddala and Kim (1998, Chapter 13).

Critical values for the ADF statistic. Under the null hypothesis of a unit root, the ADF statistic does *not* have a normal distribution, even in large samples. Because its distribution is nonstandard, the usual critical values from the normal distribution cannot be used when using the ADF statistic to test for a unit root; a special set of critical values, based on the distribution of the ADF statistic under the null hypothesis, must be used instead.

The critical values for the ADF test are given in Table 14.5. Because the alternative hypothesis of stationarity implies that $\delta < 0$ in Equations (14.32) and (14.33), the ADF test is one-sided. For example, if the regression does not include a time trend, then the hypothesis of a unit root is rejected at the 5% significance level if the ADF statistic is less than -2.86. If a time trend is included in the regression, the critical value is instead -3.41.

The critical values in Table 14.5 are substantially larger (more negative) than the one-sided critical values of -1.28 (at the 10% level) and -1.645 (at the 5% level) from the standard normal distribution. The nonstandard distribution of the ADF statistic is an example of how OLS t-statistics for regressors with stochastic trends can have nonnormal distributions. Why the large-sample distribution of the ADF statistic is nonstandard is discussed further in Section 16.3.

Does U.S. inflation have a stochastic trend? The null hypothesis that inflation has a stochastic trend can be tested against the alternative that it is stationary by performing the ADF test for a unit autoregressive root. The ADF regression with four lags of Inf_t is

$$\widehat{\Delta Inf_t} = \underset{(0.21)}{0.51} - \underset{(0.04)}{0.11} Inf_{t-1} - \underset{(0.08)}{0.19}\Delta Inf_{t-1} - \underset{(0.08)}{0.26}\Delta Inf_{t-2} + \underset{(0.08)}{0.20}\Delta Inf_{t-3} + \underset{(0.08)}{0.01}\Delta Inf_{t-4}. \tag{14.34}$$

The ADF t-statistic is the t-statistic testing the hypothesis that the coefficient on Inf_{t-1} is zero; this is $t = -2.69$. From Table 14.5, the 5% critical value is -2.86. Because the ADF statistic of -2.69 is less negative than -2.86, the test does not reject the null hypothesis at the 5% significance level. Based on the regression in Equation (14.34), we therefore cannot reject (at the 5% significance level) the null hypothesis that inflation has a unit autoregressive root, that is, that inflation contains a stochastic trend, against the alternative that it is stationary.

TABLE 14.5 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

The ADF regression in Equation (14.34) includes four lags of ΔInf_t to compute the ADF statistic. When the number of lags is estimated using the AIC, where $0 \le p \le 5$, the AIC estimator of the lag length is, however, three. When three lags are used (that is, when ΔInf_{t-1}, ΔInf_{t-2}, and ΔInf_{t-3} are included as regressors), the ADF statistic is -2.72, which is less negative than -2.86. Thus, when the number of lags in the ADF regression is chosen by AIC, the hypothesis that inflation contains a stochastic trend is not rejected at the 5% significance level.

These tests were performed at the 5% significance level. At the 10% significance level, however, the tests reject the null hypothesis of a unit root: The ADF statistics of -2.69 (four lags) and -2.72 (three lags) are more negative than the 10% critical value of -2.57. Thus the ADF statistics paint a rather ambiguous picture, and the forecaster must make an informed judgment about whether to model inflation as having a stochastic trend. Clearly, inflation in Figure 14.1a exhibits long-run swings, consistent with the stochastic trend model. In practice, many forecasters treat U.S. inflation as having a stochastic trend, and we follow that strategy here.

Avoiding the Problems Caused by Stochastic Trends

The most reliable way to handle a trend in a series is to transform the series so that it does not have the trend. If the series has a stochastic trend, that is, if the series has a unit root, then the first difference of the series does not have a trend. For example, if Y_t follows a random walk so that $Y_t = \beta_0 + Y_{t-1} + u_t$, then $\Delta Y_t = \beta_0 + u_t$ is stationary. Thus using first differences eliminates random walk trends in a series.

In practice, you can rarely be sure whether a series has a stochastic trend. Recall that, as a general point, failure to reject the null hypothesis does not necessarily mean that the null hypothesis is true; rather, it simply means that you have insufficient evidence to conclude that it is false. Thus failure to reject the null hypothesis of a unit root using the ADF test does not mean that the series actually *has* a unit root. For example, in an AR(1) model the true coefficient β_1 might be very close to 1, say 0.98, in which case the ADF test would have low power, that is, a low probability of correctly rejecting the null hypothesis in samples the size of our inflation series. Even though failure to reject the null hypothesis of a unit root does not mean the series has a unit root, it still can be reasonable to approximate the true autoregressive root as equaling 1 and therefore to use differences of the series rather than its levels.[5]

14.7 Nonstationarity II: Breaks

A second type of nonstationarity arises when the population regression function changes over the course of the sample. In economics, this can occur for a variety of reasons, such as changes in economic policy, changes in the structure of the economy, or an invention that changes a specific industry. If such changes, or "breaks," occur, then a regression model that neglects those changes can provide a misleading basis for inference and forecasting.

This section presents two strategies for checking for breaks in a time series regression function over time. The first strategy looks for potential breaks from the perspective of hypothesis testing and entails testing for changes in the regression coefficients using *F*-statistics. The second strategy looks for potential breaks from the perspective of forecasting: You pretend that your sample ends sooner than it actually does and evaluate the forecasts you would have made had this been so. Breaks are detected when the forecasting performance is substantially poorer than expected.

What Is a Break?

Breaks can arise either from a discrete change in the population regression coefficients at a distinct date or from a gradual evolution of the coefficients over a longer period of time.

One source of discrete breaks in macroeconomic data is a major change in macroeconomic policy. For example, the breakdown of the Bretton Woods system of fixed exchange rates in 1972 produced the break in the time series behavior of the $/£ exchange rate that is evident in Figure 14.2b. Prior to 1972, the exchange rate was essentially constant, with the exception of a single devaluation in 1968 in which the official value of the pound, relative to the dollar, was decreased. In contrast, since 1972 the exchange rate has fluctuated over a very wide range.

Breaks also can occur more slowly as the population regression evolves over time. For example, such changes can arise because of slow evolution of economic policy and ongoing changes in the structure of the economy. The methods for detecting breaks described in this section can detect both types of breaks, distinct changes and slow evolution.

Problems caused by breaks. If a break occurs in the population regression function during the sample, then the OLS regression estimates over the full sample will estimate a relationship that holds "on average," in the sense that the estimate combines the two different periods. Depending on the location and the size of the break, the "average" regression function can be quite different from the true regression function at the end of the sample, and this leads to poor forecasts.

Testing for Breaks

One way to detect breaks is to test for discrete changes, or breaks, in the regression coefficients. How this is done depends on whether the date of the suspected break (the **break date**) is known.

Testing for a break at a known date. In some applications you might suspect that there is a break at a known date. For example, if you are studying international trade relationships using data from the 1970s, you might hypothesize that there is a break in the population regression function of interest in 1972 when the Bretton Woods system of fixed exchange rates was abandoned in favor of floating exchange rates.

If the date of the hypothesized break in the coefficients is known, then the null hypothesis of no break can be tested using a binary variable interaction regression of the type discussed in Chapter 8 (Key Concept 8.4). To keep things simple, consider an ADL(1,1) model, so there is an intercept, a single lag of Y_t, and a single lag of X_t. Let τ denote the hypothesized break date and let $D_t(\tau)$ be a binary variable that equals 0 before the break date and 1 after, so $D_t(\tau) = 0$ if $t \leq \tau$ and $D_t(\tau) = 1$ if $t > \tau$. Then the regression including the binary break indicator and all interaction terms is

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_0 D_t(\tau) + \gamma_1[D_t(\tau) \times Y_{t-1}] + \gamma_2[D_t(\tau) \times X_{t-1}] + u_t. \tag{14.35}$$

If there is not a break, then the population regression function is the same over both parts of the sample, so the terms involving the break binary variable $D_t(\tau)$ do not enter Equation (14.35). That is, under the null hypothesis of no break, $\gamma_0 = \gamma_1 = \gamma_2 = 0$. Under the alternative hypothesis that there is a break, then the

[5] For additional discussion of stochastic trends in economic time series variables and of the problems they pose for regression analysis, see Stock and Watson (1988).

population regression function is different before and after the break date τ, in which case at least one of the γ's is nonzero. Thus the hypothesis of a break can be tested using the F-statistic that tests the hypothesis that $\gamma_0 = \gamma_1 = \gamma_2 = 0$ against the hypothesis that at least one of the γ's is nonzero. This is often called a Chow test for a break at a known break date, named for its inventor, Gregory Chow (1960).

If there are multiple predictors or more lags, then this test can be extended by constructing binary variable interaction variables for all the regressors and testing the hypothesis that all the coefficients on terms involving $D_t(\tau)$ are zero.

This approach can be modified to check for a break in a subset of the coefficients by including only the binary variable interactions for that subset of regressors of interest.

Testing for a break at an unknown break date. Often the date of a possible break is unknown or known only within a range. Suppose, for example, that you suspect that a break occurred sometime between two dates, τ_0 and τ_1. The Chow test can be modified to handle this by testing for breaks at all possible dates τ in between τ_0 and τ_1, and then using the largest of the resulting F-statistics to test for a break at an unknown date. This modified Chow test is variously called the **Quandt likelihood ratio (QLR) statistic** (Quandt, 1960) (the term we shall use) or, more obscurely, the *sup-Wald statistic*.

Because the QLR statistic is the largest of many F-statistics, its distribution is not the same as an individual F-statistic. Instead, the critical values for the QLR statistic must be obtained from a special distribution. Like the F-statistic, this distribution depends on the number of restrictions being tested, q, that is, the number of coefficients (including the intercept) that are being allowed to break, or change, under the alternative hypothesis. The distribution of the QLR statistic also depends on τ_0/T and τ_1/T, that is, on the endpoints, τ_0 and τ_1, of the subsample over which the F-statistics are computed, expressed as a fraction of the total sample size.

For the large-sample approximation to the distribution of the QLR statistic to be a good one, the subsample endpoints, τ_0 and τ_1, cannot be too close to the beginning or the end of the sample. For this reason, in practice the QLR statistic is computed over a "trimmed" range, or subset, of the sample. A common choice is to use 15% trimming, that is, to set for $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$ (rounded to the nearest integer). With 15% trimming, the F-statistic is computed for break dates in the central 70% of the sample.

The critical values for the QLR statistic, computed with 15% trimming, are given in Table 14.6. Comparing these critical values with those of the $F_{q,\infty}$ distribution (Appendix Table 4) shows that the critical values for the QLR statistics are larger. This reflects the fact that the QLR statistic looks at the largest of many individual F-statistics. By examining F-statistics at many possible break dates, the QLR statistic has many opportunities to reject the null hypothesis, leading to QLR critical values that are larger than the individual F-statistic critical values.

TABLE 14.6 Critical Values of the QLR Statistic with 15% Trimming

Number of Restrictions (q)	10%	5%	1%
1	7.12	8.68	12.16
2	5.00	5.86	7.78
3	4.09	4.71	6.02
4	3.59	4.09	5.12
5	3.26	3.66	4.53
6	3.02	3.37	4.12
7	2.84	3.15	3.82
8	2.69	2.98	3.57
9	2.58	2.84	3.38
10	2.48	2.71	3.23
11	2.40	2.62	3.09
12	2.33	2.54	2.97
13	2.27	2.46	2.87
14	2.21	2.40	2.78
15	2.16	2.34	2.71
16	2.12	2.29	2.64
17	2.08	2.25	2.58
18	2.05	2.20	2.53
19	2.01	2.17	2.48
20	1.99	2.13	2.43

These critical values apply when $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$ (rounded to the nearest integer), so the F-statistic is computed for all potential break dates in the central 70% of the sample. The number of restrictions q is the number of restrictions tested by each individual F-statistic. Critical values for other trimming percentages are given in Andrews (2003).

Like the Chow test, the QLR test can be used to focus on the possibility that there are breaks in only some of the regression coefficients. This is done by first computing the Chow tests at different break dates using binary variable interactions only for the variables with the suspect coefficients, then computing the maximum of those Chow tests over the range $\tau_0 \le \tau \le \tau_1$. The critical values for this

KEY CONCEPT 14.9

The QLR Test for Coefficient Stability

Let $F(\tau)$ denote the F-statistic testing the hypothesis of a break in the regression coefficients at date τ; in the regression in Equation (14.35), for example, this is the F-statistic testing the null hypothesis that $\gamma_0 = \gamma_1 = \gamma_2 = 0$. The QLR (or sup-Wald) test statistic is the largest of statistics in the range $\tau_0 \leq \tau \leq \tau_1$:

$$\text{QLR} = \max[F(\tau_0), F(\tau_0 + 1), \ldots, F(\tau_1)]. \quad (14.36)$$

1. Like the F-statistic, the QLR statistic can be used to test for a break in all or just some of the regression coefficients.
2. In large samples, the distribution of the QLR statistic under the null hypothesis depends on the number of restrictions being tested, q, and on the endpoints τ_0 and τ_1 as a fraction of T. Critical values are given in Table 14.6 for 15% trimming ($\tau_0 = 0.15T$ and $\tau_1 = 0.85T$, rounded to the nearest integer).
3. The QLR test can detect a single discrete break, multiple discrete breaks, and/or slow evolution of the regression function.
4. If there is a distinct break in the regression function, the date at which the largest Chow statistic occurs is an estimator of the break date.

version of the QLR test are also taken from Table 14.6, where the number of restrictions (q) is the number of restrictions tested by the constituent F-statistics.

If there is a discrete break at a date within the range tested, then the QLR statistic will reject with high probability in large samples. Moreover, the date at which the constituent F-statistic is at its maximum, $\hat{\tau}$, is an estimate of the break date τ. This estimate is a good one in the sense that, under certain technical conditions, $\hat{\tau}/T \xrightarrow{p} \tau/T$; that is, the fraction of the way through the sample at which the break occurs is estimated consistently.

The QLR statistic also rejects the null hypothesis with high probability in large samples when there are multiple discrete breaks or when the break comes in the form of a slow evolution of the regression function. This means that the QLR statistic detects forms of instability other than a single discrete break. As a result, if the QLR statistic rejects the null hypothesis, it can mean that there is a single discrete break, that there are multiple discrete breaks, or that there is slow evolution of the regression function.

The QLR statistic is summarized in Key Concept 14.9.

Warning: You probably don't know the break date even if you think you do. Sometimes an expert might believe that he or she knows the date of a possible break so that the Chow test can be used instead of the QLR test. But if this knowledge is based on the expert's knowledge of the series being analyzed, then in fact this date was estimated using the data, albeit in an informal way. Preliminary estimation of the break date means that the usual F critical values cannot be used for the Chow test for a break at that date. Thus it remains appropriate to use the QLR statistic in this circumstance.

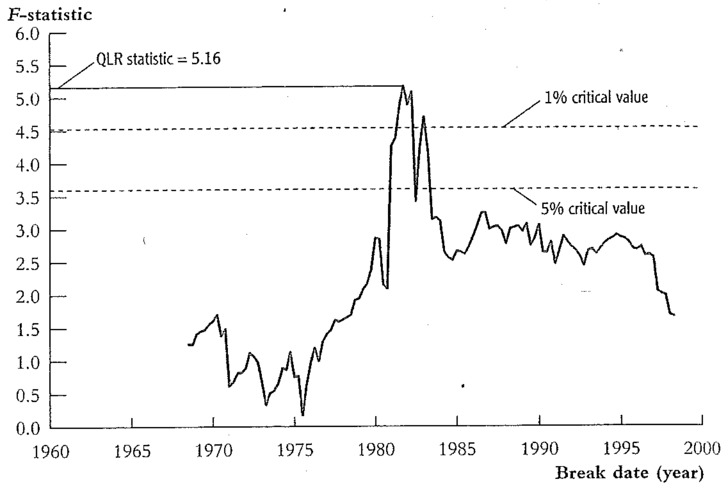
Application: Has the Phillips curve been stable? The QLR test provides a way to check whether the Phillips curve has been stable from 1962 to 2004. Specifically, we focus on whether there have been changes in the coefficients on the lagged values of the unemployment rate and the intercept in the ADL(4,4) specification in Equation (14.17) containing four lags each of ΔInf_t and $Unemp_t$.

The Chow F-statistics testing the hypothesis that the intercept and the coefficients on $Unemp_{t-1}, \ldots, Unemp_{t-4}$ in Equation (14.17) are constant against the alternative that they break at a given date are plotted in Figure 14.5 for breaks in the central 70% of the sample. For example, the F-statistic testing for a break in 1980:I is 2.85, the value plotted at that date in the figure. Each F-statistic tests five restrictions (no change in the intercept and in the four coefficients on lags of the unemployment rate), so $q = 5$. The largest of these F-statistics is 5.16, which occurs in 1981:IV; this is the QLR statistic. Comparing 5.16 to the critical values for $q = 5$ in Table 14.6 indicates that the hypothesis that these coefficients are stable is rejected at the 1% significance level (the critical value is 4.53). Thus there is evidence that at least one of these five coefficients changed over the sample.

Pseudo Out-of-Sample Forecasting

The ultimate test of a forecasting model is its out-of-sample performance, that is, its forecasting performance in "real time," after the model has been estimated. **Pseudo out-of-sample forecasting** is a method for simulating the real-time performance of a forecasting model. The idea of pseudo out-of-sample forecasting is simple: Pick a date near the end of the sample, estimate your forecasting model using data up to that date, then use that estimated model to make a forecast. Performing this exercise for multiple dates near the end of your sample yields a series of pseudo forecasts and thus pseudo forecast errors. The pseudo forecast errors can then be examined to see whether they are representative of what you would expect if the forecasting relationship were stationary.

FIGURE 14.5 *F*-Statistics Testing for a Break in Equation (14.17) at Different Dates



At a given break date, the *F*-statistic plotted here tests the null hypothesis of a break in at least one of the coefficients on $Unemp_{t-1}$, $Unemp_{t-2}$, $Unemp_{t-3}$, $Unemp_{t-4}$, or the intercept in Equation (14.17). For example, the *F*-statistic testing for a break in 1980:I is 2.85. The QLR statistic is the largest of these *F*-statistics, which is 5.16. This exceeds the 1% critical value of 4.53.

The reason this is called "pseudo" out-of-sample forecasting is that it is not true out-of-sample forecasting. True out-of-sample forecasting occurs in real time; that is, you make your forecast without the benefit of knowing the future values of the series. In pseudo out-of-sample forecasting, you simulate real-time forecasting using your model, but you have the "future" data against which to assess those simulated, or pseudo, forecasts. Pseudo out-of-sample forecasting mimics the forecasting process that would occur in real time, but without having to wait for new data to arrive.

Pseudo out-of-sample forecasting gives a forecaster a sense of how well the model has been forecasting at the end of the sample. This can provide valuable information, either bolstering confidence that the model has been forecasting well or suggesting that the model has gone off track in the recent past. The methodology of pseudo out-of-sample forecasting is summarized in Key Concept 14.10.

Other uses of pseudo out-of-sample forecasting. A second use of pseudo out-of-sample forecasting is to estimate the RMSFE. Because the pseudo out-of-sample forecasts are computed using only data prior to the forecast date, the pseudo out-of-sample forecast errors reflect both the uncertainty associated with future values of the error term and the uncertainty arising because the regression coefficients were estimated; that is, the pseudo out-of-sample forecast errors include both sources of error in Equation (14.21). Thus the sample standard deviation of the pseudo out-of-sample forecast errors is an estimator of the RMSFE. As discussed in Section 14.4, this estimator of the RMSFE can be used to quantify forecast uncertainty and to construct forecast intervals.

Pseudo Out-of-Sample Forecasts

KEY CONCEPT 14.10

Pseudo out-of-sample forecasts are computed using the following steps:

1. Choose a number of observations, P, for which you will generate pseudo out-of-sample forecasts; for example, P might be 10% or 15% of the sample size. Let $s = T - P$.
2. Estimate the forecasting regression using the shortened data set for $t = 1, \ldots, s$.
3. Compute the forecast for the first period beyond this shortened sample, $s + 1$; call this $\tilde{Y}_{s+1|s}$.
4. Compute the forecast error, $\tilde{u}_{s+1} = Y_{s+1} - \tilde{Y}_{s+1|s}$.
5. Repeat steps 2 through 4 for the remaining dates, $s = T - P + 1$ to $T - 1$ (re-estimate the regression at each date). The pseudo out-of-sample forecasts are $\{\tilde{Y}_{s+1|s}, s = T - P, \ldots, T - 1\}$, and the pseudo out-of-sample forecast errors are $\{\tilde{u}_{s+1}, s = T - P, \ldots, T - 1\}$.

A third use of pseudo out-of-sample forecasting is to compare two or more candidate forecasting models. Two models that appear to fit the data equally well can perform quite differently in a pseudo out-of-sample forecasting exercise. When the models are different, for example, when they include different predictors, pseudo out-of-sample forecasting provides a convenient way to compare the two models that focuses on their potential to provide reliable forecasts.

Application: Did the Phillips curve change during the 1990s? Using the QLR statistic, we rejected the null hypothesis that the Phillips curve has been stable against the alternative of a break at the 1% significance level (see Figure 14.5). The maximal *F*-statistic occurred in 1981:IV, indicating that a break occurred in

Can You Beat the Market? Part II

Perhaps you have heard the advice that you should buy a stock when its earnings are high relative to its price. Buying a stock is, in effect, buying the stream of future dividends paid by that company out of its earnings. If the dividend stream is unusually large relative to the price of the company's stock, then the company could be considered undervalued. If current dividends are an indicator of future dividends, then the dividend yield—the ratio of current dividends to the stock price—might forecast future excess stock returns. If the dividend yield is high, the stock is undervalued and returns would be forecasted to go up.

This reasoning suggests examining autoregressive distributed lag models of excess returns, where the predictor variable is the dividend yield. But a difficulty arises with this approach: The dividend yield is highly persistent and might even contain a stochastic trend. Using monthly data from 1960:1 to 2002:12 on the logarithm of the dividend–price ratio for the CRSP value-weighted index (the data are described in Appendix 14.1), a Dickey–Fuller unit root test including an intercept fails to reject the null hypothesis of a unit root at the 10% significance level. As always, this failure to reject the null hypothesis does not mean that the null hypothesis is true, but it does underscore that the dividend yield is a highly persistent regressor. Following the logic of Section 14.6, this result suggests that we should use the first difference of the log dividend yield as a regressor, not the level of the log dividend yield.

Table 14.7 presents ADL models of excess returns on the CRSP value-weighted index. In columns (1) and (2), the dividend yield appears in first differences, and the individual t-statistics and joint F-statistics fail to reject the null hypothesis of no predictability. But while these specifications accord with the modeling recommendations of Section 14.6, they do not correspond to the economic reasoning in the introductory paragraph, which relates returns to the *level* of the dividend yield. Column (3) of Table 14.7 therefore reports an ADL(1,1) model of excess returns using the log dividend yield, estimated through 1992:12. The t-statistic is 2.25, which exceeds the usual 5% critical value of 1.96. However, because the regressor is highly persistent, the distribution of this t-statistic is suspect and the 1.96 critical value may be inappropriate. (The F-statistic for this regression is not reported because it does not necessarily have a chi-squared distribution, even in large samples, because of the persistence of the regressor.)

One way to evaluate the apparent predictability found in column (3) of Table 14.7 is to conduct a pseudo out-of-sample forecasting analysis. Doing so over the out-of-sample period 1993:1–2002:12 provides a sample root mean square forecast error of 4.08%. In contrast, the sample RMSFEs of always forecasting excess returns to be zero is 4.00%, and the sample RMSFE of a "constant forecast" (in which the recursively estimated forecasting model includes only an intercept) is 3.98%. The pseudo out-of-sample forecast based on the ADL(1,1) model with the log dividend yield does worse than forecasts in which there are no predictors!

This lack of predictability is consistent with the strong form of the efficient markets hypothesis, which holds that all publicly available information is incorporated into stock prices so that returns should not be predictable using publicly available information (the weak form concerns forecasts based on past returns only). The core message that excess returns are not easily predicted makes sense: If they were, the prices of stocks would be driven up to the point that no expected excess returns would exist.

The interpretation of results like those in Table 14.7 is a matter of heated debate among financial economists. Some consider the lack of predictability in predictive regressions to be a vindication of the efficient markets hypothesis (see, for example, Goyal and Welch, 2003). Others say that regressions over longer time periods and longer horizons, when analyzed using tools that are specifically designed to handle persistent regressors, show evidence of predictability (see Campbell and Yogo, 2006). This predictability might arise from rational economic behavior, in which investor attitudes toward risk change over the business cycle (Campbell, 2003), or it might reflect "irrational exuberance" (Shiller, 2005).

The results in Table 14.7 concern monthly returns, but some financial econometricians have focused on ever-shorter horizons. The theory of "market microstructure"—the minute-to-minute movements of the stock market—suggests that there can be fleeting periods of predictability and that money can be made by the clever and nimble. But doing so requires nerve, plus lots of computing power—and a staff of talented econometricians.

TABLE 14.7 Autoregressive Distributed Lag Models of Monthly Excess Stock Returns

Dependent variable: excess returns on the CRSP value-weighted index.

	(1)	(2)	(3)
Specification	ADL(1,1)	ADL(2,2)	ADL(1,1)
Estimation period	1960:1– 2002:12	1960:1– 2002:12	1960:1– 1992:12
Regressors			
excess return$_{t-1}$	0.059 (0.158)	0.042 (0.162)	0.078 (0.057)
excess return$_{t-2}$		−0.213 (0.193)	
$\Delta\ln(\textit{dividend yield}_{t-1})$	0.009 (0.157)	−0.012 (0.163)	
$\Delta\ln(\textit{dividend yield}_{t-2})$		−0.161 (0.185)	
$\ln(\textit{dividend yield}_{t-1})$			0.026[a] (0.012)
Intercept	0.0031 (0.0020)	0.0037 (0.0021)	0.090[a] (0.039)
F-statistic on all coefficients (p-value)	0.501 (0.606)	0.843 (0.497)	
$\overline{R}^2$	−0.0014	−0.0008	0.0134

Notes: The data are described in Appendix 14.1. Entries in the regressor rows are coefficients, with standard errors in parentheses. The final two rows report the F-statistic testing the hypothesis that all the coefficients in the regression are zero, with its p-value in parentheses, and the adjusted R^2.

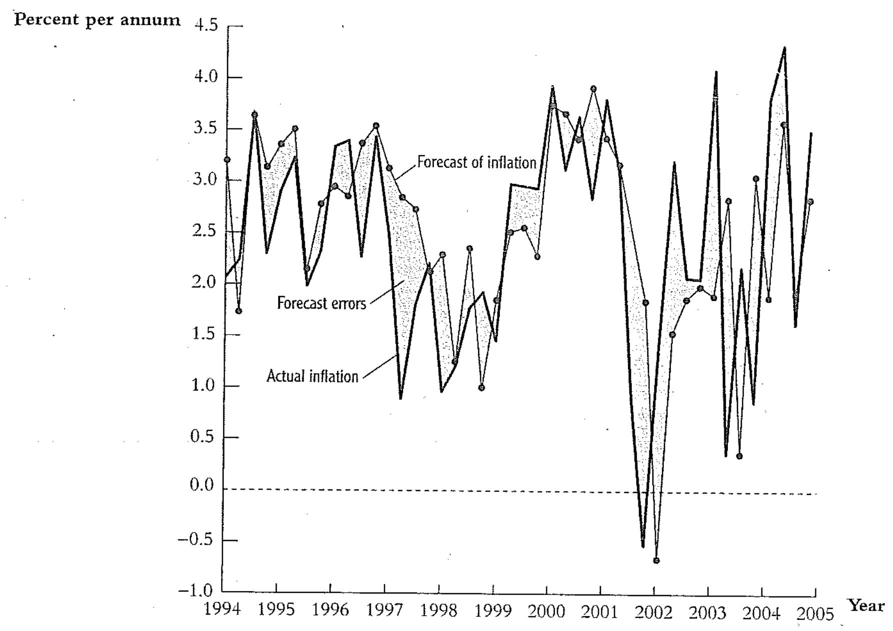
[a] $|t|>1.96$.

the early 1980s. This suggests that a forecaster using lagged unemployment to forecast inflation should use an estimation sample starting after the break in 1981:IV. Even so, a question remains: Does the Phillips curve provide a stable forecasting model subsequent to the 1981:IV break?

If the coefficients of the Phillips curve changed some time during the 1982:I–2004:I period, then pseudo out-of-sample forecasts computed using data starting in 1982:I should deteriorate. The pseudo out-of-sample forecasts of inflation for the period 1999:I–2004:IV, computed using the four-lag Phillips curve estimated with data starting 1982:I, are plotted in Figure 14.6 along with the actual values of inflation. For example, the forecast of inflation for 1999:I was computed by regressing ΔInf_t on $\Delta Inf_{t-1}, \dots, \Delta Inf_{t-4}$, $Unemp_{t-1}, \dots, Unemp_{t-4}$ with an intercept using the data through 1998:IV, then computing the forecast $\widehat{\Delta Inf}_{1999:I|1998:IV}$ using these estimated coefficients and the data through 1998:IV. The inflation forecast for 1999:I is then $\widehat{\Delta Inf}_{1999:I|1998:IV} = Inf_{1998:IV} + \widehat{\Delta Inf}_{1999:I|1998:IV}$. This entire procedure was repeated using data through 1999:I to compute the forecast $\widehat{\Delta Inf}_{1999:II|1999:I}$. Doing this for all 24 quarters from 1999:I to 2004:IV creates 24 pseudo out-of-sample forecasts, which are plotted in Figure 14.6. The pseudo out-of-sample forecast errors are the differences between actual inflation and its pseudo out-of-sample forecast, that is, the differences between the two lines in Figure 14.6. For example, in 2000:IV, the inflation rate fell by 0.8 percentage point, but the pseudo out-of-sample forecast of $\Delta Inf_{2000:IV}$ was 0.3 percentage point, so the pseudo out-of-sample forecast error was $\Delta Inf_{2000:IV} - \widehat{\Delta Inf}_{2000:IV|2000:III} = -0.8 - 0.3 = -1.1$ percentage points. In other words, a forecaster using the ADL(4,4) model of the Phillips curve, estimated through 2000:III, would have forecasted that inflation would increase by 0.3 percentage point in 2000:IV, whereas in reality it fell by 0.8 percentage point.

How do the mean and standard deviation of the pseudo out-of-sample forecast errors compare with the in-sample fit of the model? The standard error of the regression of the four-lag Phillips curve fit using data from 1982:I through 1998:IV is 1.30, so based on the in-sample fit we would expect the out-of-sample forecast errors to have mean zero and root mean square forecast error of 1.30. In fact, over the 1999:I–2004:IV pseudo out-of-sample forecast period, the average forecast error is 0.11 and the t-statistic testing the hypothesis that the mean forecast error equals zero is 0.41; thus the hypothesis that the forecasts have mean zero is not rejected. In addition, the RMSFE over the pseudo out-of-sample forecast period is 1.32, very close to value of 1.30 for the standard error of the regression for the 1982:I–1998:IV period. Moreover, the plot of the forecasts and the forecast errors in Figure 14.6 shows no major outliers or unusual discrepancies.

FIGURE 14.6 U.S. Inflation and Pseudo Out-of-Sample Forecasts



The pseudo out-of-sample forecasts made using a four-lag Phillips curve of the form in Equation (14.17) generally track actual inflation and are consistent with a stable post-1982 Phillips curve forecasting model.

According to the pseudo out-of-sample forecasting exercise, the performance of the Phillips curve forecasting model during the pseudo out-of-sample period of 1999:I–2004:IV was comparable to its performance during the in-sample period of 1982:I–1998:IV. Although the QLR test points to instability in the Phillips curve in the early 1980s, this pseudo out-of-sample analysis suggests that, after the early 1980s break, the Phillips curve forecasting model has been stable.

Avoiding the Problems Caused by Breaks

The best way to adjust for a break in the population regression function depends on the source of that break. If a distinct break occurs at a specific date, this break will be detected with high probability by the QLR statistic, and the break date can be estimated. Thus the regression function can be estimated using a binary variable indicating the two subsamples associated with this break, interacted with the other regressors as needed. If all the coefficients break, then this regression takes the form of Equation (14.35), where τ is replaced by the estimated break date, $\hat{\tau}$, while if only some of the coefficients break, then only the relevant interaction terms appear in the regression. If there is in fact a distinct break, then inference on the regression coefficients can proceed as usual, for example, using the usual normal critical values for hypothesis tests based on t-statistics. In addition, forecasts can be produced using the estimated regression function that applies to the end of the sample.

If the break is not distinct but rather arises from a slow, ongoing change in the parameters, the remedy is more difficult and goes beyond the scope of this book.[6]

14.8 Conclusion

In time series data, a variable generally is correlated from one observation, or date, to the next. A consequence of this correlation is that linear regression can be used to forecast future values of a time series based on its current and past values. The starting point for time series regression is an autoregression, in which the regressors are lagged values of the dependent variable. If additional predictors are available, then their lags can be added to the regression.

This chapter has considered several technical issues that arise when estimating and using regressions with time series data. One such issue is determining the number of lags to include in the regressions. As discussed in Section 14.5, if the number of lags is chosen to minimize the BIC, then the estimated lag length is consistent for the true lag length.

Another of these issues concerns whether the series being analyzed are stationary. If the series are stationary, then the usual methods of statistical inference (such as comparing t-statistics to normal critical values) can be used, and because the population regression function is stable over time, regressions estimated using historical data can be used reliably for forecasting. If, however, the series are nonstationary, then things become more complicated, where the specific complication depends on the nature of the nonstationarity. For example, if the series is nonstationary because it has a stochastic trend, then the OLS estimator and t-statistic can have nonstandard (nonnormal) distributions, even in large samples, and forecast performance can be improved by specifying the regression in first differences. A test for detecting this type of nonstationarity—the augmented Dickey–Fuller test for a unit root—was introduced in Section 14.6. Alternatively, if the population regression function has a break, then neglecting this break results in estimating an average version of the population regression function that in turn can lead to biased and/or imprecise forecasts. Procedures for detecting a break in the population regression function were introduced in Section 14.7.

In this chapter, the methods of time series regression were applied to economic forecasting, and the coefficients in these forecasting models were not given a causal interpretation. You do not need a causal relationship to forecast, and ignoring causal interpretations liberates the quest for good forecasts. In some applications, however, the task is not to develop a forecasting model but rather to estimate causal relationships among time series variables, that is, to estimate the *dynamic* causal effect on Y *over time* of a change in X. Under the right conditions, the methods of this chapter, or closely related methods, can be used to estimate dynamic causal effects, and that is the topic of the next chapter.

[6]For additional discussion of estimation and testing in the presence of discrete breaks, see Hansen (2001). For an advanced discussion of estimation and forecasting when there are slowly evolving coefficients, see Hamilton (1994, Chapter 13).

Summary

1. Regression models used for forecasting need not have a causal interpretation.
2. A time series variable generally is correlated with one or more of its lagged values; that is, it is serially correlated.
3. An autoregression of order p is a linear multiple regression model in which the regressors are the first p lags of the dependent variable. The coefficients of an AR(p) can be estimated by OLS, and the estimated regression function can be used for forecasting. The lag order p can be estimated using an information criterion such as the BIC.
4. Adding other variables and their lags to an autoregression can improve forecasting performance. Under the least squares assumptions for time series regression (Key Concept 14.6), the OLS estimators have normal distributions in large samples and statistical inference proceeds the same way as for cross-sectional data.
5. Forecast intervals are one way to quantify forecast uncertainty. If the errors are normally distributed, an approximate 68% forecast interval can be constructed as the forecast plus or minus an estimate of the root mean squared forecast error.

6. A series that contains a stochastic trend is nonstationary, violating the second least squares assumption in Key Concept 14.6. The OLS estimator and t-statistic for the coefficient of a regressor with a stochastic trend can have a nonstandard distribution, potentially leading to biased estimators, inefficient forecasts, and misleading inferences. The ADF statistic can be used to test for a stochastic trend. A random walk stochastic trend can be eliminated by using first differences of the series.
7. If the population regression function changes over time, then OLS estimates neglecting this instability are unreliable for statistical inference or forecasting. The QLR statistic can be used to test for a break, and, if a discrete break is found, the regression function can be re-estimated in a way that allows for the break.
8. Pseudo out-of-sample forecasts can be used to assess model stability toward the end of the sample, to estimate the root mean squared forecast error, and to compare different forecasting models.

Key Terms

first lag (520)
j^{th} lag (520)
first difference (520)
autocorrelation (522)
serial correlation (522)
autocorrelation coefficient (522)
j^{th} autocovariance (522)
autoregression (526)
forecast error (527)
root mean squared forecast error (RMSFE) (528)
p^{th} order autoregressive model [AR(p)] (529)
autoregressive distributed lag (ADL) model (534)
ADL(p,q) (534)
stationarity (535)
weak dependence (537)
Granger causality statistic (538)
Granger causality test (538)
forecast interval (540)
Bayes information criterion (BIC) (543)
Akaike information criterion (AIC) (543)
trend (546)
deterministic trend (546)
stochastic trend (546)
random walk (547)
random walk with drift (548)
unit root (549)
spurious regression (550)
Dickey–Fuller test (551)
Dickey–Fuller statistic (552)
augmented Dickey–Fuller (ADF) statistic (552)
break date (557)
Quandt likelihood ratio (QLR) statistic (558)
pseudo out-of-sample forecasting (561)

Review the Concepts

14.1 Look at the plot of the logarithm of GDP for Japan in Figure 14.2c. Does this time series appear to be stationary? Explain. Suppose that you calculated the first difference of this series. Would it appear to be stationary? Explain.

14.2 Many financial economists believe that the random walk model is a good description of the logarithm of stock prices. It implies that the percentage changes in stock prices are unforecastable. A financial analyst claims to have a new model that makes better predictions than the random walk model. Explain how you would examine the analyst's claim that his model is superior.

14.3 A researcher estimates an AR(1) with an intercept and finds that the OLS estimate of β_1 is 0.95, with a standard error of 0.02. Does a 95% confidence interval include $\beta_1 = 1$? Explain.

14.4 Suppose that you suspected that the intercept in Equation (14.17) changed in 1992:I. How would you modify the equation to incorporate this change? How would you test for a change in the intercept? How would you test for a change in the intercept if you did not know the date of the change?

Exercises

14.1 Consider the AR(1) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$. Suppose that the process is stationary.

a. Show that $E(Y_t) = E(Y_{t-1})$. (*Hint:* Read Key Concept 14.5.)

b. Show that $E(Y_t) = \beta_0/(1 - \beta_1)$.

14.2 The index of industrial production (IP_t) is a monthly time series that measures the quantity of industrial commodities produced in a given month. This problem uses data on this index for the United States. All regressions are estimated over the sample period 1960:1 to 2000:12 (that is, January 1960 through December 2000). Let $Y_t = 1200 \times \ln(IP_t/IP_{t-1})$.

a. The forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

b. Suppose that a forecaster estimates the following AR(4) model for Y_t:

$$\underset{(0.062)}{\hat{Y}_t = 1.377} + \underset{(0.078)}{0.318Y_{t-1}} + \underset{(0.055)}{0.123Y_{t-2}} + \underset{(0.068)}{0.068Y_{t-3}} + \underset{(0.056)}{0.001Y_{t-4}}.$$

Use this AR(4) to forecast the value of Y_t in January 2001 using the following values of *IP* for August 2000 through December 2000:

Date	2000:7	2000:8	2000:9	2000:10	2000:11	2000:12
IP	147.595	148.650	148.973	148.660	148.206	147.300

c. Worried about potential seasonal fluctuations in production, the forecaster adds Y_{t-12} to the autoregression. The estimated coefficient on Y_{t-12} is -0.054 with a standard error of 0.053. Is this coefficient statistically significant?

d. Worried about a potential break, she computes a QLR test (with 15% trimming) on the constant and AR coefficients in the AR(4) model. The resulting QLR statistic was 3.45. Is there evidence of a break? Explain.

e. Worried that she might have included too few or too many lags in the model, the forecaster estimates AR(p) models for $p = 1, \ldots, 6$ over the same sample period. The sum of squared residuals from each of these estimated models is shown in the table. Use the BIC to estimate the number of lags that should be included in the autoregression. Do the results differ if you use the AIC?

AR Order	1	2	3	4	5	6
SSR	29,175	28,538	28,393	28,391	28,378	28,317

14.3 Using the same data as in Exercise 14.2, a researcher tests for a stochastic trend in $\ln(IP_t)$ using the following regression:

$$\widehat{\Delta\ln(IP_t)} = \underset{(0.024)}{0.061} + \underset{(0.00001)}{0.00004t} - \underset{(0.007)}{0.018\ln(IP_{t-1})} + \underset{(0.075)}{0.333\Delta\ln(IP_{t-1})} + \underset{(0.055)}{0.162\Delta\ln(IP_{t-2})}$$

where the standard errors shown in parentheses are computed using the homoskedasticity-only formula and the regressor "t" is a linear time trend.

a. Use the ADF statistic to test for a stochastic trend (unit root) in $\ln(IP)$.

b. Do these results support the specification used in Exercise 14.2? Explain.

14.4 The forecaster in Exercise 14.2 augments her AR(4) model for *IP* growth to include four lagged values of ΔR_t ,where R_t is the interest rate on three-month U.S. Treasury bills (measured in percentage points at an annual rate).

a. The Granger-causality F-statistic on the four lags of ΔR_t is 2.35. Do interest rates help to predict IP growth? Explain.

b. The researcher also regresses ΔR_t on a constant, four lags of ΔR_t and four lags of *IP* growth. The resulting Granger-causality F-statistic on the four lags of *IP* growth is 2.87. Does IP growth help to predict interest rates? Explain.

14.5 Prove the following results about conditional means, forecasts, and forecast errors:

a. Let W be a random variable with mean μ_W and variance σ_W^2 and let c be a constant. Show that $E[(W-c)^2] = \sigma_W^2 + (\mu_W - c)^2$.

b. Consider the problem of forecasting Y_t using data on $Y_{t-1}, Y_{t-2}, \ldots$. Let f_{t-1} denote some forecast of Y_t, where the subscript $t-1$ on f_{t-1} indicates that the forecast is a function of data through date $t-1$. Let $E[(Y_t - f_{t-1})^2 | Y_{t-1}, Y_{t-2}, \ldots]$ be the conditional mean squared error of the forecast f_{t-1}, conditional on Y observed through date $t-1$. Show that the conditional mean squared forecast error is minimized when $f_{t-1} = Y_{t|t-1}$, where $Y_{t|t-1} = E(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$. (*Hint:* Review Exercise 2.27.)

c. Let u_t denote the error in Equation (14.14). Show that $\text{cov}(u_t, u_{t-j}) = 0$ for $j \neq 0$. [*Hint:* Use Equation (2.27).]

14.6 In this exercise you will conduct a Monte Carlo experiment that studies the phenomenon of spurious regression discussed in Section 14.6. In a Monte Carlo study, artificial data are generated using a computer, and then these artificial data are used to calculate the statistics being studied. This makes it possible to compute the distribution of statistics for known models when mathematical expressions for those distributions are complicated (as they are here) or even unknown. In this exercise, you will generate data so that two series, Y_t and X_t, are independently distributed random walks. The specific steps are as follows:

i. Use your computer to generate a sequence of $T = 100$ i.i.d. standard normal random variables. Call these variables $e_1, e_2, \ldots, e_{100}$. Set $Y_1 = e_1$ and $Y_t = Y_{t-1} + e_t$ for $t = 2, 3, \ldots, 100$.

ii. Use your computer to generate a new sequence, $a_1, a_2, \ldots, a_{100}$, of $T = 100$ i.i.d. standard normal random variables. Set $X_1 = a_1$ and $X_t = X_{t-1} + a_t$ for $t = 2, 3, \ldots, 100$.

iii. Regress Y_t onto a constant and X_t. Compute the OLS estimator, the regression R^2, and the (homoskedastic-only) t-statistic testing the null hypothesis that β_1 (the coefficient on X_t) is zero.

Use this algorithm to answer the following questions:

a. Run the algorithm (i) through (iii) once. Use the t-statistic from (iii) to test the null hypothesis that $\beta_1 = 0$ using the usual 5% critical value of 1.96. What is the R^2 of your regression?

b. Repeat (a) 1000 times, saving each value of R^2 and the t-statistic. Construct a histogram of the R^2 and t-statistic. What are the 5%, 50%, and 95% percentiles of the distributions of the R^2 and the t-statistic? In what fraction of your 1000 simulated data sets does the t-statistic exceed 1.96 in absolute value?

c. Repeat (b) for different numbers of observations, for example, $T = 50$ and $T = 200$. As the sample size increases, does the fraction of times that you reject the null hypothesis approach 5%, as it should because you have generated Y and X to be independently distributed? Does this fraction seem to approach some other limit as T gets large? What is that limit?

14.7 Suppose that Y_t follows the stationary AR(1) model $Y_t = 2.5 + 0.7Y_{t-1} + u_t$, where u_t is i.i.d. with $E(u_t) = 0$ and $\text{var}(u_t) = 9$.

a. Compute the mean and variance of Y_t. (*Hint:* See Exercise 14.1.)

b. Compute the first two autocovariances of Y_t. (*Hint:* Read Appendix 14.2.)

c. Compute the first two autocorrelations of Y_t.

d. Suppose that $Y_T = 102.3$. Compute $Y_{T+1|T} = E(Y_{T+1}|Y_T, Y_{t-1}, \ldots)$.

14.8 Suppose that Y_t is the monthly value of the number of new home construction projects started in the United States. Because of the weather, Y_t has a pronounced seasonal pattern; for example, housing starts are low in January and high in June. Let μ_{Jan} denote the average value of housing starts in January and $\mu_{Feb}, \mu_{Mar}, \ldots, \mu_{Dec}$ denote the average values in the other months. Show that the values of $\mu_{Jan}, \mu_{Feb}, \ldots, \mu_{Dec}$ can be estimated from the OLS regression $Y_t = \beta_0 + \beta_1 Feb_t + \beta_2 Mar_t + \cdots + \beta_{11} Dec_t + u_t$, where Feb_t is a binary variable equal to 1 if t is February, Mar_t is a binary variable equal to 1 if t is March, and so forth. Show that $\beta_0 + \beta_2 = \mu_{Mar}$, and so forth.

14.9 The moving average model of order q has the form

$$Y_t = \beta_0 + e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q},$$

where e_t is a serially uncorrelated random variable with mean 0 and variance σ_e^2.

a. Show that $E(Y_t) = \beta_0$.

b. Show that the variance of Y_t is $\text{var}(Y_t) = \sigma_e^2(1 + b_1^2 + b_2^2 + \cdots + b_q^2)$.

c. Show that $\rho_j = 0$ for $j > q$.

d. Suppose that $q = 1$. Derive the autocovariances for Y.

14.10 A researcher carries out a QLR test using 25% trimming, and there are $q = 5$ restrictions. Answer the following questions using the values in Table 14.6 ("Critical Values of the QLR Statistic with 15% Trimming") and Appendix Table 4 ("Critical Values of the $F_{m,\infty}$ Distribution").

a. The QLR F-statistic is 4.2. Should the researcher reject the null hypothesis at the 5% level?

b. The QLR F-statistic is 2.1. Should the researcher reject the null hypothesis at the 5% level?

c. The QLR F-statistic is 3.5. Should the researcher reject the null hypothesis at the 5% level?

14.11 Suppose that ΔY_t follows the AR(1) model $\Delta Y_t = \beta_0 + \beta_1 \Delta Y_{t-1} + u_t$.

a. Show that Y_t follows an AR(2) model.

b. Derive the AR(2) coefficients for Y_t as a function of β_0 and β_1.

Empirical Exercises

On the textbook Web site **www.pearsonhighered.com/stock_watson,** you will find a data file **USMacro_Quarterly** that contains quarterly data on several macroeconomic series for the United States; the data are described in the file **USMacro_Description.** Compute $Y_t = \ln(GDP_t)$, the logarithm of real GDP, and ΔY_t, the quarterly growth rate of GDP. In Empirical Exercises 14.1

through 14.6, use the sample period 1955:1–2009:4 (where data before 1955 may be used, as necessary, as initial values for lags in regressions).

E14.1 **a.** Estimate the mean of ΔY_t.

b. Express the mean growth rate in percentage points at an annual rate. [*Hint:* Multiply the sample mean in (a) by 400.]

c. Estimate the standard deviation of ΔY_t. Express your answer in percentage points at an annual rate.

d. Estimate the first four autocorrelations of ΔY_t. What are the units of the autocorrelations (quarterly rates of growth, percentage points at an annual rate, or no units at all)?

E14.2 **a.** Estimate an AR(1) model for ΔY_t. What is the estimated AR(1) coefficient? Is the coefficient statistically significantly different from zero? Construct a 95% confidence interval for the population AR(1) coefficient.

b. Estimate an AR(2) model for ΔY_t. Is the AR(2) coefficient statistically significantly different from zero? Is this model preferred to the AR(1) model?

c. Estimate AR(3) and AR(4) models. (*i*) Using the estimated AR(1) through AR(4) models, use BIC to choose the number of lags in the AR model. (*ii*) How many lags does AIC choose?

E14.3 Use an augmented Dickey–Fuller statistic to test for a unit autogressive root in the AR model for Y_t. As an alternative, suppose that Y_t is stationary around a deterministic trend.

E14.4 Test for a break in the AR(1) model for ΔY_t using a QLR test.

E14.5 **a.** Let R_t denote the interest rate for three-month treasury bills. Estimate an ADL(1,4) model for ΔY_t using lags of ΔR_t as additional predictors. Comparing the ADL(1,4) model to the AR(1) model, by how much has the $\bar{R}^2$ changed?

b. Is the Granger causality *F*-statistic significant?

c. Test for a break in the coefficients on the constant term and coefficients on the lagged values of ΔR using a QLR test. Is there evidence of a break?

E14.6 **a.** Construct pseudo out-of-sample forecasts using the AR(1) model beginning in 1989:4 and going through the end of the sample. (That is, compute $\widehat{\Delta Y}_{1990:1|1989:4}$, $\widehat{\Delta Y}_{1990:2|1990:1}$, and so forth.)

b. Construct pseudo out-of-sample forecasts using the ADL(1,4) model.

c. Construct pseudo out-of-sample using the following "naive" model:

$$\Delta Y_{t+1/t} = (\Delta Y_t + \Delta Y_{t-1} + \Delta Y_{t-2} + \Delta Y_{t-3})/4.$$

d. Compute the pseudo out-of-sample forecast errors for each model. Are any of the forecasts biased? Which model has the smallest root mean squared forecast error (RMSFE)? How large is the RMSFE (expressed in percentage points at an annual rate) for the best model?

E14.7 Read the boxes "Can You Beat the Market? Part I" and "Can You Beat the Market? Part II" in this chapter. Next, go to the course Web site, where you will find an extended version of the data set described in the boxes; the data are in the file **Stock_Returns_1931_2002** and are described in the file **Stock_Returns_1931_2002_Description**.

a. Repeat the calculations reported in Table 14.3 using regressions estimated over the 1932:1–2002:12 sample period.

b. Repeat the calculations reported in Table 14.7 using regressions estimated over the 1932:1–2002:12 sample period.

c. Is the variable ln(*dividend yield*) highly persistent? Explain.

d. Construct pseudo out-of-sample forecasts of excess returns over the 1983:1–2002:12 period using regressions that begin in 1932:1.

e. Do the results in (a) through (d) suggest any important changes to the conclusions reached in the boxes? Explain.

APPENDIX

14.1 Time Series Data Used in Chapter 14

Macroeconomic time series data for the United States are collected and published by various government agencies. The U.S. Consumer Price Index is measured using monthly surveys and is compiled by the Bureau of Labor Statistics (BLS). The unemployment rate is computed from the BLS's Current Population Survey (see Appendix 3.1). The quarterly data used here were computed by averaging the monthly values. The federal funds rate data are the monthly average of daily rates as reported by the Federal Reserve, and the dollar/pound exchange rate data are the monthly average of daily rates; both are for the final month in the quarter. Japanese GDP data were obtained from the OECD. The daily percentage change in the NYSE Composite Index was computed as $100\Delta\ln(NYSE_t)$, where $NYSE_t$ is the value of the index at the daily close of the New York Stock Exchange; because the stock exchange is not open on weekends and holidays, the time period of analysis is a business day. These and thousands of other economic time series are freely available on the Web sites maintained by various data-collecting agencies.

The regressions in Tables 14.3 and 14.7 use monthly financial data for the United States. Stock prices (P_t) are measured by the broad-based (NYSE and AMEX) value-weighted index of stock prices constructed by the Center for Research in Security Prices (CRSP). The monthly percent excess return is $100 \times \{\ln[(P_t + Div_t)/P_{t-1}] - \ln(TBill_t)\}$, where Div_t is the dividends paid on the stocks in the CRSP index and $TBill_t$ is the gross return (1 plus the interest rate) on a 30-day Treasury bill during month t. The dividend–price ratio is constructed as the dividends over the past 12 months, divided by the price in the current month. We thank Motohiro Yogo for his help and for providing these data.

APPENDIX

14.2 Stationarity in the AR(1) Model

This appendix shows that if $|\beta_1| < 1$ and u_t is stationary, then Y_t is stationary. Recall from Key Concept 14.5 that the time series variable Y_t is stationary if the joint distribution of $(Y_{s+1}, \ldots, Y_{s+T})$ does not depend on s regardless of the value of T. To streamline the argument, we show this formally for $T = 2$ under the simplifying assumptions that $\beta_0 = 0$ and $\{u_t\}$ are i.i.d. $N(0, \sigma_u^2)$.

The first step is deriving an expression for Y_t in terms of the u_t's. Because $\beta_0 = 0$, Equation (14.8) implies that $Y_t = \beta_1 Y_{t-1} + u_t$. Substituting $Y_{t-1} = \beta_1 Y_{t-2} + u_{t-1}$ into this expression yields $Y_t = \beta_1(\beta_1 Y_{t-2} + u_{t-1}) + u_t = \beta_1^2 Y_{t-2} + \beta_1 u_{t-1} + u_t$. Continuing this substitution another step yields $Y_t = \beta_1^3 Y_{t-3} + \beta_1^2 u_{t-2} + \beta_1 u_{t-1} + u_t$, and continuing indefinitely yields

$$Y_t = u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \beta_1^3 u_{t-3} + \ldots = \sum_{i=0}^{\infty} \beta_1^i u_{t-i}. \qquad (14.37)$$

Thus Y_t is a weighted average of current and past u_t's. Because the u_t's are normally distributed and because the weighted average of normal random variables is normal (Section 2.4), Y_{s+1} and Y_{s+2} have a bivariate normal distribution. Recall from Section 2.4 that the bivariate normal distribution is completely determined by the means of the two variables, their variances, and their covariance. Thus, to show that Y_t is stationary, we need to show that the means, variances, and covariance of (Y_{s+1}, Y_{s+2}) do not depend on s. An extension of the argument used below can be used to show that the distribution of $(Y_{s+1}, Y_{s+2}, \ldots, Y_{s+T})$ does not depend on s.

The means and variances of Y_{s+1} and Y_{s+2} can be computed using Equation (14.37), with the subscript $s + 1$ or $s + 2$ replacing t. First, because $E(u_t) = 0$ for all t, $E(Y_t) = E(\Sigma_{i=0}^{\infty}\beta_1^i u_{t-i}) = \Sigma_{i=0}^{\infty}\beta_1^i E(u_{t-i}) = 0$, so the mean of Y_{s+1} and Y_{s+2} are both zero and in particular do not depend on s. Second, $\text{var}(Y_t) = \text{var}(\Sigma_{i=0}^{\infty}\beta_1^i u_{t-i}) = \Sigma_{i=0}^{\infty}(\beta_1^i)^2 \text{var}(u_{t-i}) = \sigma_u^2 \Sigma_{i=0}^{\infty}(\beta_1^i)^2 = \sigma_u^2/(1 - \beta_1^2)$, where the final equality follows from the fact that if $|a| < 1$, $\Sigma_{i=0}^{\infty} a^i = 1/(1 - a)$; thus $\text{var}(Y_{s+1}) = \text{var}(Y_{s+2}) = \sigma_u^2/(1 - \beta_1^2)$, which does not depend on s as long as $|\beta_1| < 1$. Finally, because $Y_{s+2} = \beta_1 Y_{s+1} + u_{s+2}$, $\text{cov}(Y_{s+1}, Y_{s+2}) = E(Y_{s+1}Y_{s+2}) = E[Y_{s+1}(\beta_1 Y_{s+1} + u_{s+2})] = \beta_1 \text{var}(Y_{s+1}) + \text{cov}(Y_{s+1}, u_{s+2}) = \beta_1 \text{var}(Y_{s+1}) = \beta_1 \sigma_u^2/(1 - \beta_1^2)$. The covariance does not depend on s, so Y_{s+1} and Y_{s+2} have a joint probability distribution that does not depend on s; that is, their joint distribution is stationary. If $|\beta_1| \geq 1$, this calculation breaks down because the infinite sum in Equation (14.37) does not converge and the variance of Y_t is infinite. Thus Y_t is stationary if $|\beta_1| < 1$, but not if $|\beta_1| \geq 1$.

The preceding argument was made under the assumptions that $\beta_0 = 0$ and u_t is normally distributed. If $\beta_0 \neq 0$, the argument is similar except that the means of Y_{s+1} and Y_{s+2} are $\beta_0/(1 - \beta_1)$ and Equation (14.37) must be modified for this nonzero mean. The assumption that u_t is i.i.d. normal can be replaced with the assumption that u_t is stationary with a finite variance because, by Equation (14.37), Y_t can still be expressed as a function of current and past u_t's, so the distribution of Y_t is stationary as long as the distribution of u_t is stationary and the infinite sum expression in Equation (14.37) is meaningful in the sense that it converges, which requires that $|\beta_1| < 1$.

APPENDIX

14.3 Lag Operator Notation

The notation in this and the next two chapters is streamlined considerably by adopting what is known as lag operator notation. Let L denote the **lag operator**, which has the property that it transforms a variable into its lag. That is, the lag operator L has the property $\mathrm{L}Y_t = Y_{t-1}$. By applying the lag operator twice, one obtains the second lag: $\mathrm{L}^2 Y_t = \mathrm{L}(\mathrm{L}Y_t) = \mathrm{L}Y_{t-1} = Y_{t-2}$. More generally, by applying the lag operator j times, one obtains the j^{th} lag. In summary, the lag operator has the property that

$$\mathrm{L}Y_t = Y_{t-1},\ \mathrm{L}^2 Y_t = Y_{t-2}, \text{ and } \mathrm{L}^j Y_t = Y_{t-j}. \qquad (14.38)$$

The lag operator notation permits us to define the **lag polynomial**, which is a polynomial in the lag operator:

$$a(\mathrm{L}) = a_0 + a_1\mathrm{L} + a_2\mathrm{L}^2 + \cdots + a_p\mathrm{L}_p = \sum_{j=0}^{p} a_j \mathrm{L}^j, \qquad (14.39)$$

where $a_0, \ldots, a_p$ are the coefficients of the lag polynomial and $L^0 = 1$. The degree of the lag polynomial $a(L)$ in Equation (14.39) is p. Multiplying Y_t by $a(L)$ yields

$$a(L)Y_t = \left(\sum_{j=0}^{p} a_j L^j\right)Y_t = \sum_{j=0}^{p} a_j(L^jY_t) = \sum_{j=0}^{p} a_jY_{t-j} = a_0Y_t + a_1Y_{t-1} + \cdots + a_pY_{t-p}. \quad (14.40)$$

The expression in Equation (14.40) implies that the AR(p) model in Equation (14.14) can be written compactly as

$$a(L)Y_t = \beta_0 + u_t, \quad (14.41)$$

where $a_0 = 1$ and $a_j = -\beta_j$, for $j = 1, \ldots, p$. Similarly, an ADL(p,q) model can be written

$$a(L)Y_t = \beta_0 + c(L)X_{t-1} + u_t, \quad (14.42)$$

where $a(L)$ is a lag polynomial of degree p (with $a_0 = 1$) and $c(L)$ is a lag polynomial of degree $q - 1$.

APPENDIX 14.4 ARMA Models

The **autoregressive–moving average (ARMA) model** extends the autoregressive model by modeling u_t as serially correlated, specifically as being a distributed lag (or "moving average") of another unobserved error term. In the lag operator notation of Appendix 14.3, let $u_t = b(L)e_t$, where $b(L)$ is a lag polynomial of degree q with $b_0 = 1$ and e_t is a serially uncorrelated, unobserved random variable. Then the ARMA(p,q) model is

$$a(L)Y_t = \beta_0 + b(L)e_t, \quad (14.43)$$

where $a(L)$ is a lag polynomial of degree p with $a_0 = 1$.

Both AR and ARMA models can be thought of as ways to approximate the autocovariances of Y_t. The reason for this is that any stationary time series Y_t with a finite variance can be written either as an AR or as a MA with a serially uncorrelated error term, although the AR or MA models might need to have an infinite order. The second of these results, that a stationary process can be written in moving average form, is known as the Wold decomposition theorem and is one of the fundamental results underlying the theory of stationary time series analysis.

As a theoretical matter, the families of AR, MA, and ARMA models are equally rich, as long as the lag polynomials have a sufficiently high degree. Still, in some cases the autocovariances can be better approximated using an ARMA(p,q) model with small p and q than by a pure AR model with only a few lags. As a practical matter, however, the estimation of ARMA models is more difficult than the estimation of AR models, and ARMA models are more difficult to extend to additional regressors than are AR models.

APPENDIX 14.5 Consistency of the BIC Lag Length Estimator

This appendix summarizes the argument that the BIC estimator of the lag length, $\hat{p}$, in an autoregression is correct in large samples; that is, $\Pr(\hat{p} = p) \longrightarrow 1$. This is not true for the AIC estimator, which can overestimate p even in large samples.

BIC

First consider the special case that the BIC is used to choose among autoregressions with zero, one, or two lags, when the true lag length is one. It is shown below that (i) $\Pr(\hat{p} = 0) \longrightarrow 0$ and (ii) $\Pr(\hat{p} = 2) \longrightarrow 0$, from which it follows that $\Pr(\hat{p} = 1) \longrightarrow 1$. The extension of this argument to the general case of searching over $0 \leq p \leq p_{max}$ entails showing that $\Pr(\hat{p} < p) \longrightarrow 0$ and $\Pr(\hat{p} > p) \longrightarrow 0$; the strategy for showing these is the same as used in (i) and (ii) below.

Proof of (i) and (ii)

Proof of (i). To choose $\hat{p} = 0$ it must be the case that BIC(0) < BIC(1); that is, BIC(0) − BIC(1) < 0. Now $\text{BIC}(0) - \text{BIC}(1) = [\ln(SSR(0)/T) + (\ln T)/T] - [\ln(SSR(1)/T) + 2(\ln T)/T] = \ln(SSR(0)/T) - \ln(SSR(1)/T) - (\ln T)/T$. Now $\text{SSR}(0)/T = [(T-1)/T]s_Y^2 \xrightarrow{p} \sigma_Y^2$, $SSR(1)/T \xrightarrow{p} \sigma_u^2$, and $(\ln T)/T \longrightarrow 0$; putting these pieces together, $\text{BIC}(0) - \text{BIC}(1) \xrightarrow{p} \ln\sigma_Y^2 - \ln\sigma_u^2 > 0$ because $\sigma_Y^2 > \sigma_u^2$. It follows that $\Pr[\text{BIC}(0) < \text{BIC}(1)] \longrightarrow 0$, so $\Pr(\hat{p} = 0) \longrightarrow 0$.

Proof of (ii). To choose $\hat{p} = 2$ it must be the case that BIC(2) < BIC(1) or BIC(2) − BIC(1) < 0. Now $T[\text{BIC}(2) - \text{BIC}(1)] = T\{[\ln(SSR(2)/T) + 3(\ln T)/T] - [\ln(SSR(1)/T) + 2(\ln T)/T]\} = T\ln[SSR(2)/SSR(1)] + \ln T = -T\ln[1 + F/(T-2)] + \ln T$, where $F =$

$[SSR(1) - SSR(2)]/[SSR(2)/(T-2)]$ is the homoskedasticity-only F-statistic (Equation 7.13) testing the null hypothesis that $\beta_2 = 0$ in the AR(2). If u_t is homoskedastic,then F has a χ_1^2 asymptotic distribution; if not, it has some other asymptotic distribution. Thus pr[BIC(2) – BIC(1) $< 0] = \Pr\{T[\text{BIC}(2) - \text{BIC}(1)] < 0\} = \Pr\{-T\ln[1 + F/(T-2)] + (\ln T) < 0\} = \Pr\{T\ln[1 + F/(T-2)] > \ln T\}$. As T increases, $T\ln[1 + F/(T-2)] - F \xrightarrow{p} 0$ [a consequence of the logarithmic approximation $\ln(1 + a) \cong a$, which becomes exact as $a \longrightarrow 0$]. Thus $\Pr[\text{BIC}(2) - \text{BIC}(1) < 0] \longrightarrow \Pr(F > \ln T) \longrightarrow 0$, so $\Pr(\hat{p} = 2) \longrightarrow 0$.

AIC

In the special case of an AR(1) when zero, one, or two lags are considered, (i) applies to the AIC where the term $\ln T$ is replaced by 2, so $\Pr(\hat{p} = 0) \longrightarrow 0$. All the steps in the proof of (ii) for the BIC also apply to the AIC, with the modification that $\ln T$ is replaced by 2; thus $\Pr[\text{AIC}(2) - \text{AIC}(1) < 0] \longrightarrow \Pr(F > 2) > 0$. If u_t is homoskedastic, then $\Pr(F > 2) \longrightarrow \Pr(\chi_1^2 > 2) = 0.16$, so $\Pr(\hat{p} = 2) \longrightarrow 0.16$. In general, when $\hat{p}$ is chosen using the AIC, $\Pr(\hat{p} < p) \longrightarrow 0$ but $\Pr(\hat{p} > p)$ tends to a positive number, so $\Pr(\hat{p} = p)$ does not tend to 1.

CHAPTER 15 Estimation of Dynamic Causal Effects

In the 1983 movie *Trading Places*, the characters played by Dan Aykroyd and Eddie Murphy used inside information on how well Florida oranges had fared over the winter to make millions in the orange juice concentrate futures market, a market for contracts to buy or sell large quantities of orange juice concentrate at a specified price on a future date. In real life, traders in orange juice futures in fact do pay close attention to the weather in Florida: Freezes in Florida kill Florida oranges, the source of almost all frozen orange juice concentrate made in the United States, so its supply falls and the price rises. But precisely how much does the price rise when the weather in Florida turns sour? Does the price rise all at once, or are there delays; if so, for how long? These are questions that real-life traders in orange juice futures need to answer if they want to succeed.

This chapter takes up the problem of estimating the effect on Y now and in the future of a change in X, that is, the **dynamic causal effect** on Y of a change in X. What, for example, is the effect on the path of orange juice prices over time of a freezing spell in Florida? The starting point for modeling and estimating dynamic causal effects is the so-called distributed lag regression model, in which Y_t is expressed as a function of current and past values of X_t. Section 15.1 introduces the distributed lag model in the context of estimating the effect of cold weather in Florida on the price of orange juice concentrate over time. Section 15.2 takes a closer look at what, precisely, is meant by a dynamic causal effect.

One way to estimate dynamic causal effects is to estimate the coefficients of the distributed lag regression model using OLS. As discussed in Section 15.3, this estimator is consistent if the regression error has a conditional mean of zero given current and past values of X, a condition that (as in Chapter 12) is referred to as exogeneity. Because the omitted determinants of Y_t are correlated over time—that is, because they are serially correlated—the error term in the distributed lag model can be serially correlated. This possibility in turn requires "heteroskedasticity- and autocorrelation-consistent" (HAC) standard errors, the topic of Section 15.4.

A second way to estimate dynamic causal effects, discussed in Section 15.5, is to model the serial correlation in the error term as an autoregression and then to use this autoregressive model to derive an autoregressive distributed lag (ADL) model. Alternatively, the coefficients of the original distributed lag model can be estimated