CHAPTER 12

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 of 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 answer using time series data. But time series data pose special challenges, and overcoming those challenges requires some new techniques.

Chapters 12–14 introduce techniques for the econometric analysis of tim series data and apply these techniques to the problems of forecasting and estimating dynamic causal effects. Chapter 12 introduces the basic concepts and tools of regression with time series data and applies them to economic forecasting. In Chapter 13, the concepts and tools developed in Chapter 12 are applied to the problem of estimating dynamic causal effects using time series data. Chapter 14 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 12.1, 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 12.2 introduces some basic concepts of time series analysis and presents some examples of economic time series data. Section 12.3 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 12.4. 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 12.5 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 12.6 and 12.7.

12.1 Using Regression Models for Forecasting

Chapter 4 began by considering the problem of a school superintenden
wants to know how much test scores would increase if she reduces class si
her school district; that is, the superintendent wants to know the causal eff
test scores of a change in class size. Accordingly, Parts II and III focused on
regression analysis to estimate causal effects using observational data.

Now consider a different problem, that of a parent moving to a metr
itan area and choosing a town within that area in part based on the local s
system. The parent would like to know how the different school district
form on standardized tests. Suppose, however, that test score data are un
able (perhaps they are confidential) but data on class sizes are. The paren
must guess at how well the different districts perform on standardized tests
on a limited amount of information. That is, the parent's problem is to fo
average test scores in a given district based on information related to test s
such as class size.

The superintendent's problem and the parent's problem are conceptuall
different. Multiple regression is a powerful tool for both, but because the
lems are different, the criteria used to assess the suitability of a particular r
sion model is different as well. To obtain the credible estimates of causal
desired by the superintendent, we must worry about the issues raised in Cl
7: omitted variable bias, selection, simultaneous causality, and so forth. In
trast, to obtain the reliable forecast desired by the parent, it is important th
estimated regression have good explanatory power, that its coefficients b
mated precisely, and that it is stable in the sense that the regression estimat
one set of data can be reliably used to make forecasts using other data.

For example, recall the regression of test scores on the student-teache
(*STR*) from Chapter 4:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR.$$

We concluded that this regression is not useful for the superintenden
OLS estimator of the slope is biased because of omitted variables such as the
position of the student body and their other learning opportunities outside sc
The superintendent cannot change the district's average income or the fr
of non-English speakers, both of which affect test scores. Because these var
are also correlated with class size, there is omitted variable bias. Thus the regr

of test scores on the student-teacher ratio yields a biased estimator of the effect on test scores of a change in the student-teacher ratio, and Equation (12.1) does not answer the superintendent's question.

Nevertheless, Equation (12.1) could be useful to the parent trying to choose a district. To be sure, class size is not the only determinant of test performance, but from the parent's perspective what matters is whether it is a reliable predictor of test performance. The parent interested in forecasting test scores does not care whether the coefficient in Equation (12.1) estimates the causal effect on test scores of class size. Rather, the parent simply wants the regression to explain much of the variation in test scores across districts and to be stable, that is, to apply to the districts to which the parent is considering moving. Although omitted variable bias makes Equation (12.1) useless for answering the causal question, it still can be useful for forecasting.

The applications in this chapter are different than the test score/class size prediction problem because this chapter focuses on using time series data to forecast future events. Yet time series forecasting is similar conceptually to the parent's problem: the task is to use the known values of some variables (current and past values of the rate of price inflation instead of class size) to forecast the value of another variable (future inflation instead of test scores). As in the parent's problem, regression models can produce reliable forecasts, even if their coefficients have no causal interpretation. In Chapter 13, we return to problems like that faced by the school superintendent and discuss the estimation of causal effects using time series variables.

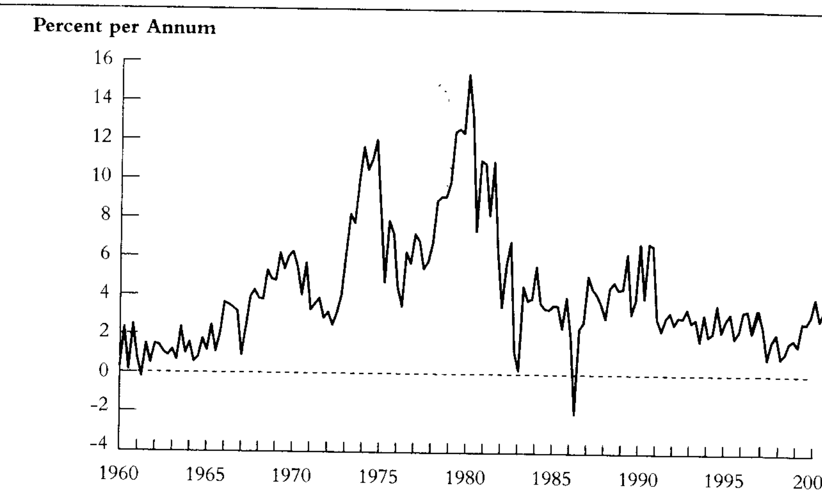
12.2 Introduction to Time Series Data and Serial Correlation

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.

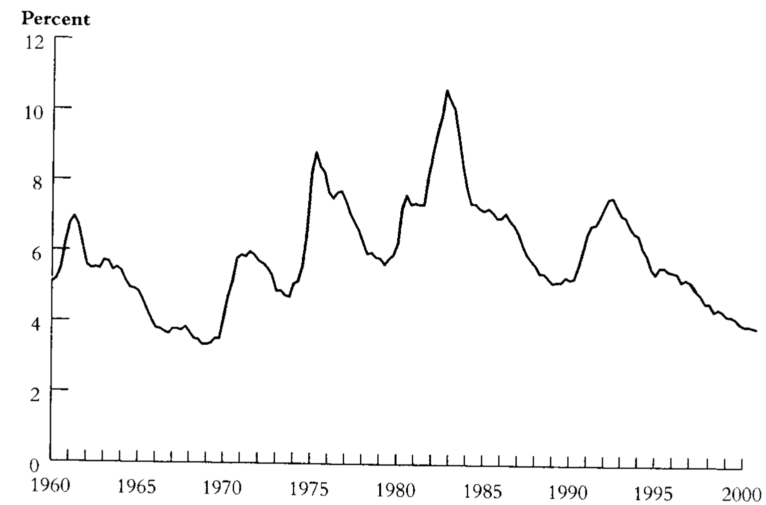
The Rates of Inflation and Unemployment in the United States

Figure 12.1a 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 1999 (the data are described in Appendix 12.1). The inflation rate

FIGURE 12.1 Inflation and Unemployment in the United States, 1960–1999



(a) U.S. CPI Inflation Rate



(b) U.S. Unemployment Rate

Price inflation in the United States (Figure 12.1a) drifted upwards from 1960 until 1980, and then fell sharply the early 1980s. The unemployment rate in the United States (Figure 12.1b) rises during recessions (the shaded episodes) and falls during expansions.

was low in the 1960s, rose through the 1970s to a postwar 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 12.1a, 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 12.1b. 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 1960–61, 1970, 1974–75, the twin recessions of 1980 and 1981–82, and the recession of 1990–91, episodes denoted by shading in Figure 12.1b.

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 **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 that $\Delta Y_t = Y_t - Y_{t-1}$.

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

Key Concept 12.1

Lags, First Differences, Logarithms, and Growth Rates

- 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.

proportional (or percentage) changes in the original series, and this is achieved taking the logarithm of the series.[1]

Lags, first differences, and growth rates are summarized in Key Concept 1

Lags, changes, and percentage changes are illustrated using the U.S. inflat rate in Table 12.1. The first column shows the date, or period, where the first qu ter of 1999 is denoted 1999:I, the second quarter of 1999 is denoted 1999:II, so forth. The second column shows the value of the CPI in that quarter, and third column shows the rate of inflation. For example, from the first to the sec quarter of 1999, the index increased from 164.9 to 166.0, a percentage increase $100 \times (166.03 - 164.87)/164.87 = 0.704\%$. This is the percentage increase from quarter to the next. It is conventional to report rates of inflation (and other gro rates in macroeconomic time series) on an annual basis, which is the percent increase in prices that would occur over a year, if the series were to continue increase at the same rate. Because there are four quarters a year, the annualized of inflation in 1999:II is $0.704 \times 4 = 2.82$, or 2.8% per year after rounding.

This percentage change can also be computed using the differences- logarithms approximation in Key Concept 12.1. The difference in the logarit of the CPI from 1999:I to 1999:II is $\ln(166.03) - \ln(164.87) = 0.00701$, yield

[1]Recall from Section 6.2 that the change of the logarithm of a variable is approximately equ the proportional change of that variable; that is, $\ln(X + a) - \ln(X) \cong a/X$, where the approxima works best when a/X is small. Now, replace X with Y_{t-1}, a with ΔY_t, and note that $Y_t = Y_{t-1} +$ This means that the proportional change in the series Y_t between periods $t-1$ and t is appr mately $\ln(Y_t) - \ln(Y_{t-1}) = \ln(Y_{t-1} + \Delta Y_t) - \ln(Y_{t-1}) \cong \Delta Y_t/Y_{t-1}$. The expression $\ln(Y_t) - \ln(Y_{t-1})$ is 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 fractional change, so the percentage change in the series Y_t is approximately $100\Delta\ln(Y_t)$.

TABLE 12.1 Inflation in the United States in 1999 and the First Quarter of 2000

Quarter	U.S. CPI	Rate of Inflation at an Annual Rate (Inf_t)	First Lag (Inf_{t-1})	Change in Inflation (ΔInf_t)
1999:I	164.87	1.6	2.0	−0.4
1999:II	166.03	2.8	1.6	1.2
1999:III	167.20	2.8	2.8	0.0
1999:IV	168.53	3.2	2.8	0.4
2000:I	170.27	4.1	3.2	0.9

The annualized rate of inflation is the percentage change in the CPI from the previous quarter to the current quarter, times 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.

the approximate quarterly percentage difference $100 \times 0.00701 = 0.701\%$. On an annualized basis, this is $0.701 \times 4 = 2.80$, or 2.8% after rounding, the same as obtained by directly computing the percentage growth. These calculations can be summarized as

$$\text{annualized rate of inflation} = Inf_t \cong 400[\ln(CPI_t) - \ln(CPI_{t-1})] = 400\Delta\ln(CPI_t), \tag{12.2}$$

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 12.1 illustrate lags and changes. The first lag of inflation in 1999:II is 1.6%, the inflation rate in 1999:I. The change in the rate of inflation from 1999:I to 1999:II was 2.8% − 1.6% = 1.2%.

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

Key Conc 12.2

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,

$$j^{th} \text{ autocovariance} = \text{cov}(Y_t, Y_{t-j}) \tag{12.3}$$

$$j^{th} \text{ autocorrelation} = \rho_j = \text{corr}(Y_t, Y_{t-j}) = \frac{\text{cov}(Y_t, Y_{t-j})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-j})}}. \tag{12.4}$$

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

correlation between Y_t and Y_{t-j}. Similarly, the ***j*^th^ autocovariance** is the c ance between Y_t and Y_{t-j}. Autocorrelation and autocovariance are summari Key Concept 12.2.

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

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

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

where $\overline{Y}_{j+1,T}$ denotes the sample average of Y_t computed over the observ $t = j + 1, \ldots, T$ and where $\widehat{\text{var}(Y_t)}$ is the sample variance of Y. (Eq (12.6) uses the assumption that $\text{var}(Y_t)$ and $\text{var}(Y_{t-j})$ are the same, an impli of the assumption that Y is stationary, which is discussed in Section 12.4.)

The first four sample autocorrelations of the inflation rate and of the c in the inflation rate are listed in Table 12.2. These entries show that infla strongly positively autocorrelated: the first autocorrelation is 0.85. The s autocorrelation declines as the lag increases, but it remains large even at a four quarters. The change in inflation is negatively autocorrelated: an incre

TABLE 12.2 First Four Sample Autocorrelations of the U.S. Inflation Rate and Its Change, 1960:I–1999:IV

	Autocorrelation of:	
Lag	**Inflation Rate (Inf_t)**	**Change of Inflation Rate (ΔInf_t)**
1	0.85	−0.24
2	0.77	−0.27
3	0.77	0.32
4	0.68	−0.06

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 12.1: 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.

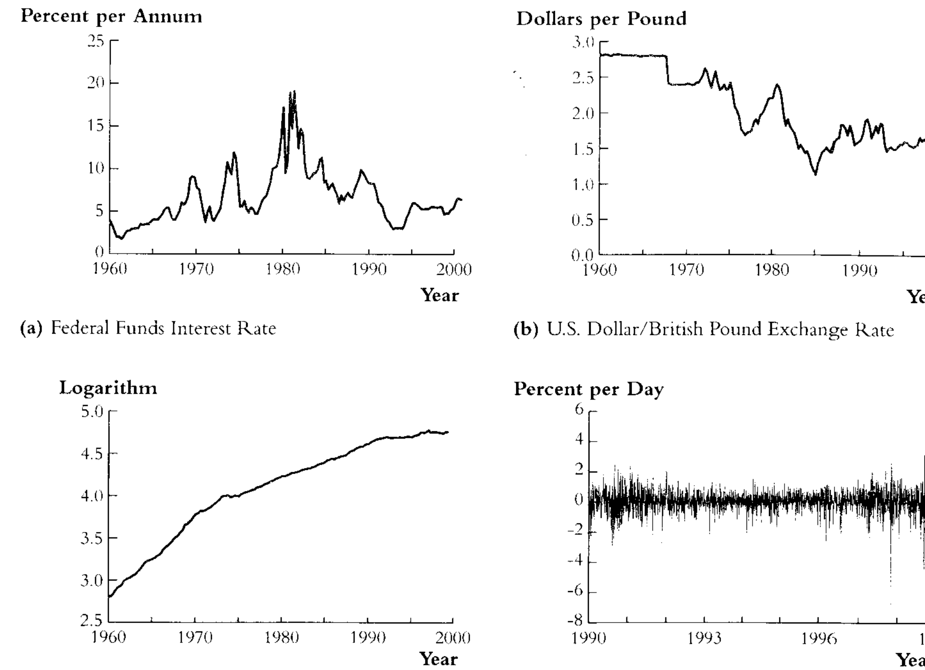
Other Examples of Economic Time Series

Economic time series differ greatly. Four examples of economic time series are plotted in Figure 12.2: the U.S. Federal Funds interest rate; the rate of exchange between the dollar and the British pound; the logarithm of real 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 12.2a) 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 12.1, you will see that sharp increases in the Federal Funds rate often have been associated with subsequent recessions.

The dollar/pound exchange rate (Figure 12.2b) 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

FIGURE 12.2 Four Economic Time Series



(a) Federal Funds Interest Rate

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

(c) Logarithm of Real GDP in Japan

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

The four time series have markedly different patterns. The Federal Funds Rate (Figure 12.2a) has a pattern simil price inflation. The exchange rate between the U.S. dollar and the British pound (Figure 12.2b) shows a discret change after the 1972 collapse of the Bretton Woods system of fixed exchange rates. The logarithm of real GDI Japan (Figure 12.2c) shows relatively smooth growth, although the growth rate decreases in the 1970s and ag the 1990s. The daily returns on the NYSE stock price index (Figure 12.2d) are essentially unpredictable, but its ance changes: this series shows "volatility clustering."

led to the breakdown of this system; thereafter, the major currencies were a to "float", that is, their values were determined by the supply and demand f rencies in the market for foreign exchange. Prior to 1972, the exchange r approximately constant, with the exception of a single devaluation in 1 which the official value of the pound, relative to the dollar, was decreased to Since 1972 the exchange rate has fluctuated over a very wide range.

Real quarterly Japanese GDP (Figure 12.2c) is the total value of goods and services produced in Japan during a quarter, adjusted for inflation. GDP is the broadest measure of total economic activity. The logarithm of the series is plotted in Figure 12.2c, and changes in this series can be interpreted as (decimal) growth rates. During the 1960s and early 1970s, Japanese real GDP grew quickly, but this growth slowed in the late 1970s and 1980s. Growth slowed further during the 1990s, averaging only 1.5% per year from 1990–1999.

The daily return on the NYSE index of stock prices (Figure 12.2d) is the percentage change from one trading day to the next of the NYSE Composite market index, a broad index of the share prices of all firms traded on the New York Stock Exchange. Figure 12.2d plots these daily returns for January 2, 1990, to December 31, 1998 (a total of 1,771 observations). Unlike the other series in Figure 12.2, there is very little serial correlation in these daily returns: if there were, then you could predict these returns using past daily returns and make money by buying when you expect the market to rise and selling when you expect it to fall. Although the return itself is essentially unpredictable, inspection of Figure 12.2d reveals patterns in the volatility of returns. For example, the standard deviation of returns was relatively large in 1991 and 1998, and relatively small in 1995. 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 14.5.

12.3 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 product, and local governments use inflation forecasts when they develop their budget 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 u data from 1962–1999, this regression is

$$\widehat{\Delta Inf_t} = \underset{(0.14)}{0.02} - \underset{(0.106)}{0.211}\Delta Inf_{t-1}, \quad (1$$

where, as usual, standard errors are given in parentheses under the estimated c ficients, and $\widehat{\Delta Inf_t}$ is the predicted value of ΔInf_t based on the estimated regres line. The model in Equation (12.7) is called a first order autoregression: an aut gression because it is a regression of the series onto its own lag, ΔInf_{t-1}, and order because only one lag is used as a regressor. The coefficient in Equa (12.7) is negative, so an increase in the inflation rate in one quarter is associ with a decline in the inflation rate in the next quarter.

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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t, \quad (1$$

where u_t is an error term.

Forecasts and forecast errors. Suppose you have historical data on Y you want to forecast its future value. If Y_t follows the AR(1) model in Eq tion (12.8) and if β_0 and β_1 are known, then the forecast of Y_t based on Y_{t-} $\beta_0 + \beta_1 Y_{t-1}$.

In practice, β_0 and β_1 are unknown, so forecasts must be based on estim of β_0 and β_1. We will use the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, which are construc using historical data. In general, $\hat{Y}_{t|t-1}$ will denote the forecast of Y_t based information through period $t-1$ using a model estimated with data through per $t-1$. Accordingly, the forecast based on the AR(1) model in Equation (12.8

$$\hat{Y}_{t|t-1} = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1} \quad (1$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using historical data through time $t-1$.

The **forecast error** is the mistake made by the forecast; this is the differe between the value of Y_t that actually occurred and its forecasted value based on

$$\text{forecast error} = Y_t - \hat{Y}_{t|t-1}. \quad (12$$

Forecasts vs. 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:

$$\text{RMSFE} = \sqrt{E[(Y_t - \hat{Y}_{t|t-1})^2]}. \tag{12.11}$$

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 (12.8)). The standard deviation of u_t is in turn estimated by the standard error of the regression (*SER*, see Section 5.10). 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 12.4.

Application to inflation. What is the forecast of inflation in the first quarter of 2000 (2000:I) that a forecaster would have made in 1999:IV, based on the estimated AR(1) model in Equation (12.7) (which was estimated using data through 1999:IV)? From Table 12.1, the inflation rate in 1999:IV was 3.2% (so $Inf_{1999:IV} = 3.2\%$), an increase of 0.4 percentage points from 1999:III (so $\Delta Inf_{1999:IV} = 0.4$). Plugging these values into Equation (12.7), the forecast of the change in inflation from 1999:IV to 2000:I is $\widehat{\Delta Inf}_{2000:I} = 0.02 - 0.211 \times \Delta Inf_{1999:IV} = 0.02 - 0.211 \times 0.4 =$

$-0.06 \cong -0.1$ (rounded to the nearest tenth). The predicted rate of inflation the past rate of inflation plus its predicted change:

$$\widehat{Inf}_{t|t-1} = Inf_{t-1} + \widehat{\Delta Inf}_{t|t-1}. \quad (12.12$$

Because $Inf_{1999:IV} = 3.2\%$ and the predicted change in the inflation rate fro 1999:IV to 2000:I is -0.1, the predicted rate of inflation in 2000:I is $\widehat{Inf}_{2000:I} =$ $Inf_{1999:IV} + \widehat{\Delta Inf}_{2000:I} = 3.2\% - 0.1\% = 3.1\%$. Thus, the AR(1) model forecasts th inflation will drop slightly from 3.2% in 1999:IV to 3.1% in 2000:I.

How accurate was this AR(1) forecast? From Table 12.1, the actual value o inflation in 2000:I was 4.1%, so the AR(1) forecast is low by a full percentag point; that is, the forecast error is 1.0%. The $\bar{R}^2$ of the AR(1) model in Equatio (12.7) is only 0.04, so the lagged change of inflation explains a very small fractio of the variation in inflation in the sample used to fit the autoregression. This lov $\bar{R}^2$ is consistent with the poor forecast of inflation in 2000:I produced using Equa tion (12.7). More generally, the low $\bar{R}^2$ suggests that this AR(1) model will fore cast only a small amount of the variation in the change of inflation.

The standard error of the regression in Equation (12.7) is 1.67; ignorin uncertainty arising from estimation of the coefficients, our estimate of the RMSF for forecasts based on Equation (12.7) therefore is 1.67 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 usefu information in the more distant past. One way to incorporate this information i to include additional lags in the AR(1) model; this yields the p^{th} order autore gressive, or AR(p), model.

The **p^{th} order autoregressive model** (the **AR(p)** model) represents Y_t as linear function of p of its lagged values; that is, in the AR(p) model, the regres sors are $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$, plus an intercept. The number of lags, p, included i 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 th change in inflation as regressors. Estimated by OLS over the period 1962–1999 the AR(4) model is

$$\underset{}{\widehat{\Delta Inf}_t} = \underset{(0.12)}{0.02} - \underset{(0.10)}{0.21}\Delta Inf_{t-1} - \underset{(0.09)}{0.32}\Delta Inf_{t-2} + \underset{(0.09)}{0.19}\Delta Inf_{t-3} - \underset{(0.10)}{0.04}\Delta Inf_{t-4}. \quad (12.13$$

The coefficients on the final three additional lags in Equation (12.13) are jointl significantly different from zero at the 5% significance level: the F-statistic is 6.4

Key Concept 12.3

Autoregressions

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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t, \qquad (12.14)$$

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.

(p-value < 0.001). This is reflected in an improvement in the $\bar{R}^2$ from 0.04 for the AR(1) model in Equation (12.7) to 0.21 for the AR(4). Similarly, the *SER* of the AR(4) model in Equation (12.13) is 1.53, an improvement over the *SER* of the AR(1) model, which is 1.67.

The AR(p) model is summarized in Key Concept 12.3.

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 based on its entire history depends on only the most recent p past values. Specifically, let $Y_{t|t-1} = E(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$ denote the conditional mean of Y_t given its entire history. Then $Y_{t|t-1}$ has the smallest RMSFE of any forecast based on the history of Y (Exercise 12.5). If Y_t follows an AR(p), then its conditional mean is

$$Y_{t|t-1} = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p}, \qquad (12.15)$$

which follows from the AR(p) model in Equation (12.14) 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 (12.15) with estimated coefficients.

The second implication is that the errors u_t are serially uncorrelated, a result that follows from Equation (2.25) (Exercise 12.5).

Application to inflation. What is the forecast of inflation in 2000:I using data through 1999:IV, based on the AR(4) model of inflation in Equation (12.13)? To compute this forecast, substitute the values of the change of inflation in each

of the four quarters of 1999 into Equation (12.13): $\widehat{\Delta Inf}_{2000:I|1999:IV} = 0.($ $0.21\Delta Inf_{1999:IV} - 0.32\Delta Inf_{1999:III} + 0.19\Delta Inf_{1999:II} - 0.04\Delta Inf_{1999:I} = 0.02 - 0.$ $0.4 - 0.32 \times 0.0 + 0.19 \times 1.1 - 0.04 \times (-0.4) \cong 0.2$, where the 1999 value the change of inflation are taken from the final column of Table 12.1.

The corresponding forecast of inflation in 2000:I is the value of inflatio 1999:IV, plus the forecasted change, that is, 3.2% + 0.2% = 3.4%. The fore error is the actual value, 4.1%, minus the forecast, or 4.1% – 3.4% = 0.7%, slig smaller than the AR(1) forecast error of 1.0%.

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

Economic theory often suggests other variables that could help to forecast variable of interest. These other variables, or predictors, can be added tc autoregression to produce a time series regression model with multiple pre tors. When other variables and their lags are added to an autoregression, the re 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 fu decline in the rate of inflation. This negative relationship, known as the short- Phillips curve, is evident in the scatterplot of Figure 12.3, in which year-to- changes in the rate of price inflation are plotted against the rate of unemployn in the previous year. For example, in 1982 the unemployment rate averaged 9. and during the next year the rate of inflation fell by 2.9%. Overall, the corr tion in Figure 12.3 is –0.40.

The scatterplot in Figure 12.3 suggests that past values of the unemployn rate might contain information about the future course of inflation that is already contained in past changes of inflation. This conjecture is readily chec by augmenting the AR(4) model in Equation (12.13) to include the first la the unemployment rate:

$$\widehat{\Delta Inf}_t = \underset{(0.55)}{1.42} - \underset{(0.09)}{0.26}\Delta Inf_{t-1} - \underset{(0.10)}{0.40}\Delta Inf_{t-2} + \underset{(0.08)}{0.11}\Delta Inf_{t-3} - \underset{(0.10)}{0.09}\Delta Inf_{t-4} - \underset{(0.10)}{0.23}Unem \qquad (12$$

FIGURE 12.3 Scatterplot of Change in Inflation Between Year *t* and Year *t* + 1 vs. the Unemployment Rate in Year *t*

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 –0.40.

The t-statistic on $Unemp_{t-1}$ is –2.33, so this term is significant at the 5% level. The $\bar{R}^2$ of this regression is 0.22, a small improvement over the AR(4) $\bar{R}^2$ of 0.21.

The forecast of the change of inflation in 2000:I is obtained by substituting the 1999 values of the change of inflation into Equation (12.16), along with the value of the unemployment rate in 1999:IV (which is 4.1%); the resulting forecast is $\widehat{\Delta Inf}_{2000:I|1999:IV} = 0.5$. Thus the forecast of inflation in 2000:I is 3.2% + 0.5% = 3.7%, and the forecast error is 0.4%. This forecast is closer to actual 2000:I inflation than was the AR(4) forecast.

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

$$\begin{aligned}\widehat{\Delta Inf}_t &= \underset{(0.47)}{1.32} - \underset{(0.09)}{0.36}\Delta Inf_{t-1} - \underset{(0.10)}{0.34}\Delta Inf_{t-2} + \underset{(0.08)}{0.07}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4} \\ &\quad - \underset{(0.47)}{2.68}Unemp_{t-1} + \underset{(0.89)}{3.43}Unemp_{t-2} - \underset{(0.89)}{1.04}Unemp_{t-3} + \underset{(0.44)}{0.07}Unemp_{t-4}.\end{aligned} \tag{12.17}$$

The F-statistic testing the joint significance of the second through fourth lags of the unemployment rate is 4.93 (p-value = 0.003), so they are jointly significant.

The $\bar{R}^2$ of the regression in Equation (12.17) is 0.35, a solid improvem
0.22 for Equation (12.16). The F-statistic on all the unemployment co
is 8.51 (p-value < 0.001), indicating that this model represents a statisti
nificant improvement over the AR(4) model of Section 12.3 (Equation
The standard error of the regression in Equation (12.17) is 1.37, a su
improvement over the SER of 1.53 for the AR(4).

The forecasted change in inflation from 1999:IV to 2000:I using
(12.17) is computed by substituting the values of the variables into the e
The unemployment rate was 4.3% in 1999:I and 1999:II, 4.2% in 199
4.1% in 1999:IV. The forecast of the change in inflation from 1999:IV t
based on Equation (12.17), is

$$\widehat{\Delta Inf}_{2000:I|1999:IV} = 1.32 - 0.36 \times 0.4 - 0.34 \times 0.0 + 0.07 \times 1.1 - 0.0$$
$$\times (-0.4) - 2.68 \times 4.1 + 3.43 \times 4.2 - 1.04 \times 4.3 + 0.07 \times 4.3 = 0.5$$

Thus the forecast of inflation in 2000:I is 3.2% + 0.5% = 3.7%. The fo
error is small, 0.4. Adding multiple lags of the unemployment rate appe
improve inflation forecasts beyond those of an AR(4).

The autoregressive distributed lag model. The models in Equati
(12.16) and (12.17) are **autoregressive distributed lag (ADL)** mode
"autoregressive" because lagged values of the dependent variable are inc
regressors, as in an autoregression, and "distributed lag" because the reg
also includes multiple lags (a "distributed lag") of an additional predicto
general, an autoregressive distributed lag model with p lags of the deper
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 (12.16) is an ADL(4,1)
and the model in Equation (12.17) is an ADL(4,4) model.

The autoregressive distributed lag model is summarized in Key Conc
With all these regressors, the notation in Equation (12.19) is somewhat
some, and alternative optional notation, based on the so-called lag op
presented in Appendix 12.3.

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

The ADL model contains lags of the dependent variable (the autor
component) and a distributed lag of a single additional predictor, X. In

The Autoregressive Distributed Lag Model

Key Concept 12.4

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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_q X_{t-q} + u_t, \quad (12.19)$$

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

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 precise definition of stationarity, given in Key Concept 12.5, is that the 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 12.6. 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 12.6 modify the four least squares assumptions of the multiple regression model for cross-sectional data (Key Concept 5.4) for time series data.

Stationarity

Key Cor 12.5

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; 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. Stationarity requires the future to be like the past, at least in a probabilistic sense.

The first assumption is that u_t has conditional mean zero, given all th sors *and* the additional lags of the regressors beyond the lags included in th sion. This assumption extends the assumption used in the AR and ADI and implies that the best forecast of Y_t using all past values of Y and the X's by the regression in Equation (12.20).

The second least squares assumption for cross-sectional data (Key 5.4) is that $(X_{1i}, \ldots, X_{ki}, Y_i)$, $i = 1, \ldots, n$, are independently and identic tributed (i.i.d.). The second assumption for time series regression replaces assumption by a more appropriate one with two parts. Part (a) is that the drawn from a stationary distribution, so that the distribution of the data the same as its distribution in the past. This assumption is a time series ve the "identically distributed" part of the i.i.d. assumption: the cross-s requirement of each draw being identically distributed is replaced by t series requirement that the joint distribution of the variables, *including l* not change over time. In practice, many economic time series appear to stationary, which means that this assumption can fail to hold in applicatio time series variables are nonstationary, then one or more problems can time series regression: the forecast can be biased, the forecast can be in (there can be alternative forecasts based on the same data with lower vari conventional OLS-based statistical inferences (for example performing a esis test by comparing the OLS t-statistic to ± 1.96) can be misleading. I which of these problems occurs, and its remedy, depends on the source of stationarity. In Sections 12.6 and 12.7, we study the problems posed by, and solutions to two empirically important types of nonstationarity in ec time series, trends and breaks. For now, however, we simply assume that t are jointly stationary and accordingly focus on regression with stationary v

ey
oncept
!.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:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \delta_{11} X_{1t-1} + \delta_{12} X_{1t-2} + \cdots + \delta_{1q_1} X_{1t-q_1} + \cdots + \delta_{k1} X_{kt-1} + \delta_{k2} X_{kt-2} + \cdots + \delta_{kq_k} X_{kt-q_k} + u_t, \quad (12.20)$$

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. $X_{1t}, \ldots, X_{kt}$ 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 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 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 12.6, inference on the regression coefficients using OLS proceeds in the same way as it usually does using cross-sectional data.

Granger Causality Tests (Tests of Predictive Content)

Key
Concept
12.7

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 (12.20) (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.

One useful application of the F-statistic in time series forecasting is to tes[t] 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 ha[s] no predictive content corresponds to the null hypothesis that the coefficients o[n] all lags of that variable are zero. The F-statistic testing this null hypothesis is calle[d] the **Granger causality statistic**, and the associated test is called a **Grange[r] causality test** (Granger (1969)). This test is summarized in Key Concept 12.7.

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 applie[d] experimentally and we observe the subsequent effect on Y. In contrast, Grange[r] causality means that if X Granger-causes Y, then X is a useful predictor of Y, give[n] the other variables in the regression. While "Granger predictability" is a mor[e] accurate term than "Granger causality," the latter has become part of the jargo[n] 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 o[n] the OLS estimates in Equation (12.17), the F-statistic testing the null hypothesis that the coefficients on all four lags of the unemployment rate are zero is 8.5[1] ($p < 0.001$): in the jargon of Key Concept 12.7, 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 rat[e] 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

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T]. \quad (12.21)$$

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 (12.21). Thus the mean squared forecast error (MSFE) is

$$\begin{aligned} \text{MSFE} &= E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] \\ &= \sigma_u^2 + \text{var}[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\delta}_1 - \delta_1)X_T], \end{aligned} \quad (12.22)$$

and the RMSFE is the square root of the MSFE.

Estimation of the MSFE entails estimation of the two parts in Equation (12.22). The first term, σ_u^2, can be estimated by the square of the standard error of the regression, as discussed in Section 12.3. The second term requires estimating the variance of a weighted average of the regression coefficients, and methods for doing so were discussed in Section 6.1 (see the discussion following Equation (6.7)).

An alternative method for estimating the MSFE is to use the variance of pseudo out-of-sample forecasts, a procedure discussed in Section 12.7.

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

One important difference between a forecast interval and a confidence
val is that the usual formula for a 95% confidence interval (the estimator
standard errors) is justified by the central limit theorem and therefore hold
wide range of distributions of the error term. In contrast, because the fo
error in Equation (12.21) includes the future value of the error u_{T+1}, to co
a forecast interval requires either estimating the distribution of the error te
making some assumption about that distribution.

In practice, it is convenient to assume that u_{T+1} is normally distributed
Equation (12.21) and the central limit theorem applied to $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\delta}_1$ imp
the forecast error is the sum of two independent, normally distributed terms,
the forecast error is itself normally distributed with variance equaling the MS
follows that a 95% confidence interval is given by $\hat{Y}_{T+1|T} \pm 1.96\, SE(Y_{T+1} - \hat{Y}$
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
homoskedastic. If instead u_{T+1} is heteroskedastic, then one needs to dev
model of the heteroskedasticity so that the term σ_u^2 in Equation (12.22)
estimated, given the most recent values of Y and X, and methods for mo
this conditional heteroskedasticity are presented in Section 14.5.

Because of uncertainty about future events—that is, uncertainty about
95% forecast intervals can be so wide that they have limited use in decision
ing. Professional forecasters therefore often report forecast intervals that are
than 95%, for example, one standard error forecast intervals (which are 68%
cast intervals if the errors are normally distributed). Alternatively, some fo
ers report multiple forecast intervals, as is done by the economists at the B
England when they publish their inflation forecasts (see the River of Bloo
on the following page).

12.5 Lag Length Selection Using Information Criteria

The estimated inflation regressions in Sections 12.3 and 12.4 have either
four lags of the predictors. One lag makes some sense, but why four? Mor
erally, how many lags should be included in a time series regression? This s
discusses statistical methods for choosing the number of lags, first in an a
gression, then in a time series regression model with multiple predictors.

The River of Blood

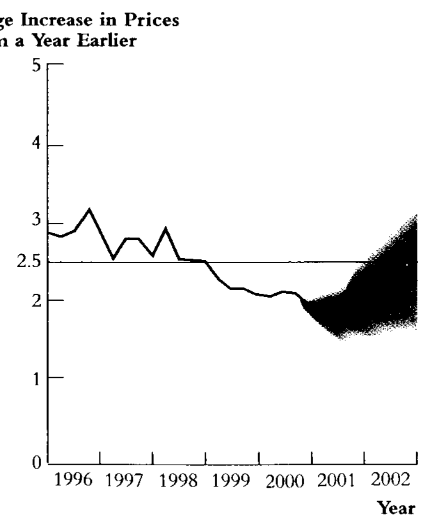
As part of its efforts to inform the public about monetary policy decisions, the Bank of England regularly publishes forecasts of inflation. These forecasts combine output from econometric models maintained by professional econometricians at the bank with the expert judgment of the members of the bank's 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 February 2001 is shown in Figure 12.4 (in this figure the blood is green, not red, so you will need to use your imagination). This chart shows that, as of February 2001, the bank's economists expected inflation to remain essentially unchanged over the next year at approximately 2%, but then to increase. There is considerable uncertainty about this forecast, however. In their written discussion, they cited in particular the possibility of a further slowdown in the United States—which in

(Continued)

FIGURE 12.4 The River of Blood

The Bank of England's fan chart for February 2001 shows forecast ranges for inflation.



fact became the recession of 2001—that could lead to lower inflation in the United Kingdom As it happened, their forecast was a good one: in the fourth quarter of 2001, the rate of inflation was 2.0%.

The Bank of England has been a pioneer in the movement towards 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 citize[ns.] In a democracy in the information age, reason[...] the economists at the Bank of England, it is part[ic]ularly important for citizens to understand [the] bank's economic outlook and the reasoning behi[nd] its difficult decisions.

To see the river of blood in its original [...] hue, visit the Bank of England's Web site **www.bankofengland.co.uk/inflationreport.**

Determining the Order of an Autoregression

In practice, choosing the order p of an autoregression requires balancin[g the] benefit of including more lags against the cost of additional estimation u[ncer]tainty. On the one hand, if the order of an estimated autoregression is too [low,] you will omit potentially valuable information contained in the more di[stant] 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 w[ith a] model with many lags and to perform hypothesis tests on the final lag. For ex[am]ple, you might start by estimating an AR(6) and test whether the coefficie[nt on] the sixth lag is significant at the 5% level; if not, drop it and estimate an A[R(5),] test the coefficient on the fifth lag, and so forth. The drawback of this meth[od 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-statisti[c 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 ti[me.]

The BIC. A way around this problem is to estimate p by minimizing an "i[nfor]mation criterion." One such information criterion is the **Bayes information [cri]terion (BIC)**, also called the **Schwarz information criterion (SIC)**, wh[ich is]

$$\mathrm{BIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}, \quad (1\ldots$$

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 $p = 0, 1, \ldots, p_{max}$, where p_{max} is the largest value of p considered.

The formula for the BIC might look a bit mysterious at first, but it has an intuitive appeal. Consider the first term in Equation (12.23). 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 12.5.

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 12.3 for autoregressions of maximum order six ($p_{max} = 6$). For example, for the AR(1) model in Equation (12.7), $SSR(1)/T = 2.726$, so $\ln(SSR(1)/T) = 1.003$. Because $T = 152$ (38 years, four quarters per year), $\ln(T)/T = 0.033$ and $(p + 1)\ln(T)/T = 2 \times 0.033 = 0.066$. Thus $\text{BIC}(1) = 1.003 + 0.066 = 1.069$.

The BIC is smallest when $p = 3$ in Table 12.3. Thus the BIC estimate of the lag length is 3. As can be seen in Table 12.3, as the number of lags increases the

TABLE 12.3 The Bayes Information Criterion (BIC) and the R^2 for Autoregressive Models of U.S. Inflation, 1962–1999

p	$SSR(p)/T$	$\ln(SSR(p)/T)$	$(p+1)\ln(T)/T$	BIC(p)	R^2
0	2.853	1.048	0.033	1.081	0.000
1	2.726	1.003	0.066	1.069	0.045
2	2.361	0.859	0.099	0.958	0.173
3	2.264	0.817	0.132	0.949	0.206
4	2.261	0.816	0.165	0.981	0.207
5	2.260	0.815	0.198	1.013	0.208
6	2.257	0.814	0.231	1.045	0.209

R^2 increases and the SSR decreases. The increase in the R^2 is large from one t two lags, smaller from two to three, and quite small from three to four. The BI helps decide precisely how large the increase in the R^2 must be to justify includ ing the additional lag.

The AIC. The BIC is not the only information criterion; another is th **Akaike information criterion,** or **AIC**:

$$\text{AIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}. \quad (12.24$$

The difference between the AIC and the BIC is that the term "$\ln T$" in th BIC is replaced by "2" in the AIC, so the second term in the AIC is smaller. Fo example, for the 152 observations used to estimate the inflation autoregression $\ln T = \ln(152) = 5.02$, so that the second term for the BIC is more than twice a large as the term in AIC. Thus a smaller decrease in the SSR is needed in the AI to justify including another lag. As a matter of theory, the second term in the AI is not large enough to ensure that the correct lag length is chosen, even in larg samples, so the AIC estimator of p is not consistent. As is discussed in Appendi 12.5, in large samples the AIC will overestimate p with nonzero probability.

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

A note on calculating information criteria. How well two estimated regres sions fit the data is best assessed when they are estimated using the same data set Because the BIC and AIC are formal methods for making this comparison, th autoregressions under consideration should be estimated using the same observa tions. For example, in Table 12.3 all the regressions were estimated using dat from 1962:I–1999:IV, for a total of 152 observations. Because the autoregressio involve lags of the change of inflation, this means that earlier values of the chang of inflation (values before 1962:I) were used as regressors for the prelimina observations. Said differently, the regressions examined in Table 12.3 each includ observations on $\Delta Inf_t, \Delta Inf_{t-1}, \ldots, \Delta Inf_{t-p}$ for $t = $ 1962:I, . . . , 1999:IV, corre sponding to 152 observations on the dependent variable and regressors, so T = 152 in Equations (12.23) and (12.24).

Lag Length Selection in Time Series Regression with Multiple Predictors

The tradeoff involved with lag length choice in the general time series regression model with multiple predictors (Equation (12.20)) 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 (12.17), we tested the hypothesis that the coefficients on the second through fourth lag 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

$$\text{BIC}(K) = \ln\left(\frac{SSR(K)}{T}\right) + K\frac{\ln T}{T}. \tag{12.25}$$

The AIC is defined in the same way, but with 2 replacing $\ln T$ in Equation (12.25). 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 (12.25), 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, t
require that $p = q_1 = \cdots = q_k$, so that only $p_{max} + 1$ models need to be c
(corresponding to $p = 0, 1, \ldots, p_{max}$).

12.6 Nonstationarity I: Trends

In Key Concept 12.6, it was assumed that the dependent variable and th
sors are stationary. If this is not the case, that is, if the dependent variabl
regressors are nonstationary, then conventional hypothesis tests, confiden
vals, and forecasts can be unreliable. The precise problem created by nons
ity, and the solution to that problem, depends on the nature of that nonsta
In this and the next section, we examine two of the most important
nonstationarity in economic time series data: trends and breaks. In each
we first describe the nature of the nonstationarity, then discuss the conse
for time series regression if this type of nonstationarity is present but is
We next present tests for nonstationarity and discuss remedies for, or s
to, the problems caused by that particular type of nonstationarity. We l
discussing trends.

What Is a Trend?

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

Inspection of Figure 12.1a suggests that the U.S. inflation rate has a tr
sisting of a general upward tendency through 1982 and a downward t
thereafter. The series in Figures 12.2a, b, and c also have trends, but the
are quite different. The trend in the U.S. Federal Funds interest rate is s
the trend in the U.S. inflation rate. The \$/£ exchange rate clearly had a p
downward trend after the collapse of the fixed exchange rate system in 1
logarithm of Japanese real GDP has a complicated trend: fast growth at f
moderate growth, and finally slow growth.

Deterministic and stochastic trends. There are two types of tre
in time series data, deterministic and stochastic. A **deterministic tr**
nonrandom function of time. For example, a deterministic trend migh
ear in time; if inflation had a deterministic linear trend so that it incr
0.1 percentage point per quarter, this trend could be written as $0.1t$, w
measured in quarters. In contrast, a **stochastic trend** is random and va

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 12.1.

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 bad monetary policy, 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 14.3.

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

$$Y_t = Y_{t-1} + u_t, \quad (12.26)$$

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 (12.26), 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 12.2c, 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 adjustme leads to an extension of the random walk model to include a tendency to mov or "drift" in one direction or the other. This extension is referred to as a **rando walk with drift**:

$$Y_t = \beta_0 + Y_{t-1} + u_t, \quad (12.2$$

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 positive, then Y_t increases on average. In the random walk with drift mod the best forecast of the series tomorrow is the value of the series today, pl the drift β_0.

The random walk model (with drift as appropriate) is simple yet versatile, a 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 n stationary: the variance of a random walk increases over time so the distributi of Y_t changes over time. One way to see this is to recognize that, because u_t serially uncorrelated in Equation (12.26), $\text{var}(Y_t) = \text{var}(Y_{t-1}) + \text{var}(u_t)$; for Y_t to stationary, $\text{var}(Y_t)$ cannot depend on time, so in particular $\text{var}(Y_t) = \text{var}(Y_{t-1})$ mu hold, but this can only happen if $\text{var}(u_t) = 0$. Another way to see this is to ima ine that Y_t starts out at zero, that is, $Y_0 = 0$. Then $Y_1 = u_1$, $Y_2 = u_1 + u_2$, and forth, so that $Y_t = u_1 + u_2 + \cdots + u_t$. Because u_t is serially uncorrelated, var($= \text{var}(u_1 + u_2 + \cdots + u_t) = t\sigma_u^2$. Thus the variance of Y_t depends on t; in fa it increases as t increases. Because the variance of Y_t depends on t, its distributi depends on t, that is, it is nonstationary.

Because the variance of a random walk increases without bound, its *popu tion* autocorrelations are not defined (the first autocovariance and variance are in nite and the ratio of the two is not well defined). However, a feature of a rando walk is that its *sample* autocorrelations tend to be very close to one, in fact, the sample autocorrelation of a random walk converges to one in probability.

Stochastic trends, autoregressive models, and a unit root. The rando walk model is a special case of the AR(1) model (Equation (12.8)) in which β_1 = In other words, if Y_t follows an AR(1) with $\beta_1 = 1$, then Y_t contains a stochas trend and is nonstationary. If, however, $|\beta_1| < 1$ and u_t is stationary, then the jo distribution of Y_t and its lags does not depend on t (a result shown in Appen 12.2) so Y_t is stationary as long as u_t is stationary.

The analogous condition for an AR(p) to be stationary is more complicat than the condition $|\beta_1| < 1$ for an AR(1). Its formal statement involves the ro 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 solutions to the equation $1 - \beta_1 z - \beta_2 z^2 - \beta_3 z^3 - \cdots - \beta_p z^p = 0$.) For an AR($p$) to be stationary, the roots of this polynomial must all be greater than one 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 one in absolute value is equivalent to $|\beta_1| < 1$.

If an AR(p) has a root that equals one, 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: first, the estimator of the autoregressive coefficient in an AR(1) is biased towards zero if its true value is one; second, t-statistics on regressors with a stochastic trends can have a nonnormal distribution, even in large samples; and third, 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 towards zero. Suppose that Y_t follows the random walk in Equation (12.26) but this is unknown to the econometrician, who instead estimates the AR(1) model in Equation (12.8). Because Y_t is nonstationary, the least squares assumptions for time series regression in Key Concept 12.6 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 towards 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 that 5% of the time $\hat{\beta}_1 < 0.824$.

One implication of this bias towards zero is that, if Y_t follows a rando
then forecasts based on the AR(1) model can perform substantially wo
those based on the random walk model, which imposes the true value β_1 =
conclusion also applies to higher order autoregressions, in which there a
casting gains from imposing a unit root (that is, from estimating the autore
in first differences instead of in levels) when in fact the series contains a u

Problem #2: Nonnormal distributions of* t*-statistics. If a regres
stochastic trend, then its usual OLS t-statistic can have a nonnormal dist
under the null hypothesis, even in large samples. This nonnormal dist
means that conventional confidence intervals are not valid and hypoth
cannot be conducted as usual. In general, the distribution of this t-statist
readily tabulated because the distribution depends on the relationship b
the regressor in question and the other regressors. One important case i
it *is* possible to tabulate this distribution is in the context of an autoregressi
a unit root, and we return to this special case when we take up the pro
testing whether a time series contains a stochastic trend.

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

For example, U.S. inflation was steadily rising from the mid-1960s
the early 1980s, and at the same time Japanese GDP was steadily rising
two trends conspire to produce a regression that appears to be "significan
conventional measures. Estimated by OLS using data from 1965 throug
this regression is

$$\widehat{U.S.Inflation}_t = -\underset{(0.08)}{2.84} + \underset{(0.02)}{0.18}Japanese\ GDP_t,\ \bar{R}^2 = 0.56.$$

The t-statistic on the slope coefficient exceeds 9, which by our usual s
indicates a strong positive relationship between the two series, and the $\bar{R}^2$
However, running this regression using data from 1982 through 1999 yi

$$\widehat{U.S.Inflation}_t = \underset{(1.37)}{6.25} - \underset{(0.01)}{0.03}Japanese\ GDP_t,\ \bar{R}^2 = 0.07.$$

The regressions in Equation (12.28) and (12.29) could hardly be m
ferent. Interpreted literally, Equation (12.28) indicates a strong positive r
ship, while Equation (12.29) indicates a weak 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 1999. 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 (12.28) and (12.29) illustrate empirically the theoretical point that OLS can be misleading when the series contain stochastic trends (see Exercise 12.6 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 14.4.

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 12.2. Because the first autocorrelation coefficient will be near one 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 14.3), 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 (12.27) 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

$$H_0: \beta_1 = 1 \text{ vs. } H_1: \beta_1 < 1 \text{ in } Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t. \quad (12.30)$$

If $\beta_1 = 1$, the AR(1) has an autoregressive root of one, so the null hypo in Equation (12.30) is that the AR(1) has a unit root, and the alternative it is stationary.

This test is most easily implemented by estimating a modified version of tion (12.30) obtained by subtracting Y_{t-1} from both sides. Let $\delta = \beta_1 - 1$ Equation (12.30) becomes

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

The OLS t-statistic testing $\delta = 0$ in Equation (12.31) is called the **Di Fuller statistic**. The formulation in Equation (12.31) is convenient be regression software automatically prints out the t-statistic testing $\delta = 0$. Not the Dickey-Fuller test is one-sided, because the relevant alternative is tha stationary so $\beta_1 < 1$ or, equivalently, $\delta < 0$. The Dickey-Fuller statistic is puted using "nonrobust" standard errors, that is, the "homoskedasticity- standard errors presented in Appendix 4.4 (Equation (4.62) for the case of gle regressor and in Section 16.4 for the multiple regression model).[2]

The Dickey-Fuller test in the AR(p) model. The Dickey-Fuller statisti sented in the context of Equation (12.31) applies only to an AR(1). As disc in Section 12.3, for some series the AR(1) model does not capture all the correlation in Y_t, in which case a higher order autoregression is more approp

The extension of the Dickey-Fuller test to the AR(p) model is summa in Key Concept 12.8. Under the null hypothesis, $\delta = 0$ and ΔY_t is a stati AR(p). Under the alternative hypothesis, $\delta < 0$ so that Y_t is stationary. Be the regression used to compute this version of the Dickey-Fuller statistic is mented by lags of ΔY_t, the resulting t-statistic is referred to as the **augme Dickey-Fuller (ADF) statistic**.

In general the lag length p is unknown, but it can be estimated using an mation criterion applied to regressions of the form (12.32) for various val p. Studies of the ADF statistic suggest that it is better to have too many lag too few, so it is recommended to use the AIC instead of the BIC to estimate the ADF statistic.[3]

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

[3]See Stock (1994) for a review of simulation studies of the finite-sample properties of the D Fuller and other unit root test statistics.

Key Concept 12.8

The Augmented Dickey-Fuller Test for a Unit Autoregressive Root

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

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t. \quad (12.32)$$

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 (12.32).

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

$$\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \cdots + \gamma_p \Delta Y_{t-p} + u_t, \quad (12.33)$$

where α is an unknown coefficient and the ADF statistic is the OLS t-statistic testing $\delta = 0$ in Equation (12.33).

The lag length p can be estimated using the BIC or AIC. 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 (12.32) or (12.33) and are given in Table 12.4.

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, like the rate of inflation, that do not exhibit long-term growth. But other economic time series, like Japanese GDP (Figure 12.2c), 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 r ified to test the null hypothesis of a unit root against the alternative that it is tionary around a linear time trend. As summarized in Equation (12.33) in Concept 12.8, this is accomplished by adding a time trend (the regressor X_t to the regression.

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

Critical values for the ADF statistic. Under the null hypothesis of a root, the ADF statistic does *not* have a normal distribution, even in large sam Because its distribution is nonstandard, the usual critical values from the no distribution cannot be used when using the ADF statistic to test for a unit a special set of critical values, based on the distribution of the ADF statistic u the null hypothesis, must be used instead.

The critical values for the ADF test are given in Table 12.4. Because the a native hypothesis of stationarity implies that $\delta < 0$ in Equations (12.32) (12.33), the ADF test is one-sided. For example, if the regression does not inc a time trend, then the hypothesis of a unit root is rejected at the 5% signific level if the ADF statistic is less than −2.86. If a time trend is included in the re sion, the critical value is instead −3.41.

The critical values in Table 12.4 are substantially larger (more negative) the one-sided critical values of −1.28 (at the 10% level) and −1.645 (at th level) from the standard normal distribution. The nonstandard distribution o ADF statistic is an example of how OLS t-statistics for regressors with stoch trends can have nonnormal distributions. Why the large-sample distributio the ADF statistic is nonstandard is discussed further in Section 14.3.

TABLE 12.4 Large-Sample Critical Values of the Augmented Dickey-Fuller Statisti

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

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

$$\widehat{\Delta Inf_t} = \underset{(0.23)}{0.53} - \underset{(0.04)}{0.11} Inf_{t-1} - \underset{(0.08)}{0.14}\Delta Inf_{t-1} - \underset{(0.08)}{0.25}\Delta Inf_{t-2} + \underset{(0.08)}{0.24}\Delta Inf_{t-3} + \underset{(0.08)}{0.01}\Delta Inf_{t-4} \quad (12.34)$$

The ADF t-statistic is the t-statistic testing the hypothesis that the coefficient on Inf_{t-1} is zero; this is $t = -2.60$. From Table 12.4, the 5% critical value is -2.86. Because the ADF statistic of -2.60 is less negative than -2.86, the test does not reject at the 5% significance level. Based on the regression in Equation (12.34), 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.

The ADF regression in Equation (12.34) includes four lags of ΔInf_t to compute the ADF statistic. When the number of lags is estimated using the AIC, where $0 \leq p \leq 6$, 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 -2.65, 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 -2.60 (four lags) and -2.65 (three lags) are slightly 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 or not to model inflation as having a stochastic trend. Clearly, inflation in Figure 12.1a exhibits long-run swings, consistent with the stochastic trend model. Moreover, 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 $Y_t = \beta_0 + Y_{t-1} + u_t$, then $\Delta Y_t = \beta$
u_t is stationary. Thus using first differences eliminates random walk trends in a se

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

12.7 Nonstationarity II: Breaks

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

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

What Is a Break?

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

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

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 12.2b. 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 than 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 or not.

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 6 (Key Concept 6.4). 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 zero before the break date and one after, so $D_t(\tau) = 0$

if $t \le \tau$ and $D_t(\tau) = 1$ if $t > \tau$. Then the regression including the binary i
cator and all interaction terms is

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

If there is not a break, then the population regression function is
over both parts of the sample so the terms involving the break binary var
do not enter Equation (12.35). That is, under the null hypothesis of
$\gamma_0 = \gamma_1 = \gamma_2 = 0$. Under the alternative hypothesis that there is a break
population regression function is different before and after the break
which case at least one of the γ's is nonzero. Thus the hypothesis of a bre
tested using the F-statistic that tests the hypothesis that $\gamma_0 = \gamma_1 = \gamma_2 = 0$ a
hypothesis that at least one of the γ's is nonzero. This is often called a C
for a break at a known break date, named for its inventor, Gregory Cho

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

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

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

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

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 version of the QLR test are also taken from Table 12.5, where the number of restrictions (q) is the number of restrictions tested by the constituent F-tests.

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 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, 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 12.9.

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

TABLE 12.5 Critical Values of the QLR Statistic with 15% Trimming

Number of Restrictions (q)	10%	5%	1
1	7.12	8.68	12
2	5.00	5.86	7
3	4.09	4.71	6
4	3.59	4.09	5.
5	3.26	3.66	4.
6	3.02	3.37	4.
7	2.84	3.15	3.
8	2.69	2.98	3.
9	2.58	2.84	3.
10	2.48	2.71	3.
11	2.40	2.62	3.
12	2.33	2.54	2.
13	2.27	2.46	2.
14	2.21	2.40	2.
15	2.16	2.34	2.
16	2.12	2.29	2.
17	2.08	2.25	2.
18	2.05	2.20	2.
19	2.01	2.17	2.
20	1.99	2.13	2.

These critical values apply when $\tau_0 = 0.15T$ and $\tau_1 = 0.85T$ (rounded to the nearest integer), so F-statistic is computed for all potential break dates in the central 70% of the sample. The nu restrictions q is the number of restrictions tested by each individual F-statistic. This table wa provided to us by Donald Andrews, and supercedes Table 1 in Andrews (1993).

in fact this date was estimated using the data, albeit in an informal way. Pr nary estimation of the break date means that the usual F critical values can used for the Chow test for a break at that date. Thus it remains appropriate the QLR statistic in this circumstance.

Key Concept 12.9: 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 (12.35), 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 is the largest of statistics in the range $\tau_0 \le \tau \le \tau_1$:

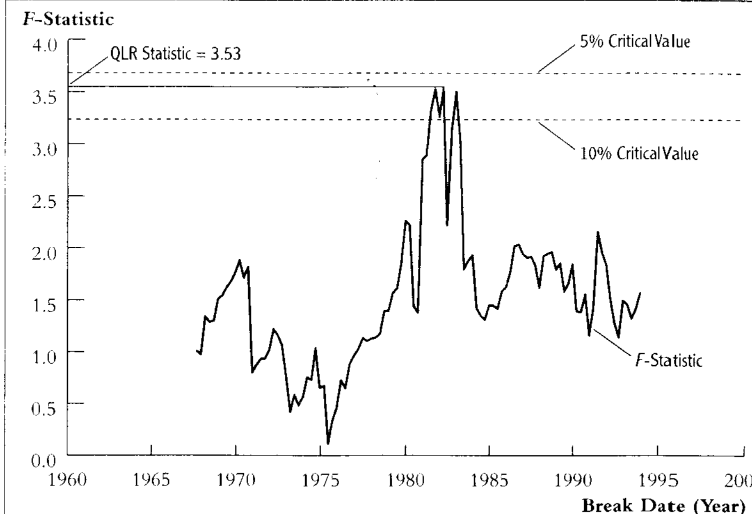
$$\text{QLR} = \max[F(\tau_0), F(\tau_0 + 1), \ldots, F(\tau_1)]. \quad (12.36)$$

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 12.5 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.

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 1999. 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 (12.17) 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 (12.17) are constant against the alternative that they break at a given date are plotted in Figure 12.5 for breaks in the central 70% of the sample. For example, the F-statistic testing for a break in 1980:I is 2.26, 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 3.53, which occurs in 1982:II; this is the QLR statistic. Comparing 3.53 to the critical values for $q = 5$ in Table 12.5 indicates that the hypothesis that these coefficients are stable is rejected at the 10% significance level (the critical value is 3.26), but not 5% significance level (the critical value is 3.66). Thus there is some evidence that at

FIGURE 12.5 ***F*-Statistics Testing for a Break in Equation (12.17) at Different Dates**



At a given break date, the F-statistic plotted here tests the null hypothesis of a break in at lea one of the coefficients on $Unemp_{t-1}$, $Unemp_{t-2}$, $Unemp_{t-3}$, $Unemp_{t-4}$, or the intercept in Equa (12.17). For example, the F-statistic testing for a break in 1980:I is 2.26. The QLR statistic is largest of these F-statistics, which is 3.53. This exceeds the 10% critical value of 3.26, but no 5% critical value of 3.66.

least one of these five coefficients has changed over the sample, but the evi is not especially strong.

Pseudo Out-of-Sample Forecasting

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

Key Concept 12.10

Pseudo Out-of-Sample Forecasts

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 2–4 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 $\{\tilde{u}_{s+1}, s = T - P, \ldots, T - 1\}$.

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 12.10.

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

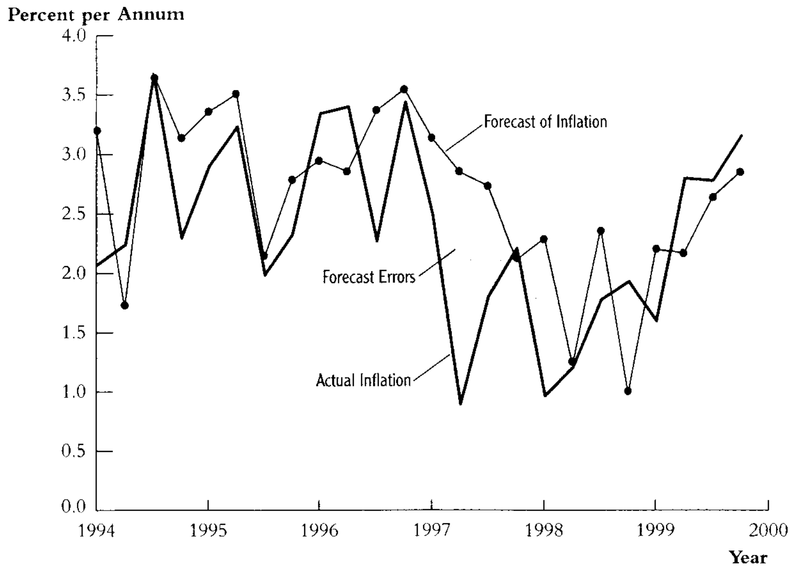
coefficients were estimated; that is, the pseudo out-of-sample forec
include both sources of error in Equation (12.21). Thus the sample stanc
ation of the pseudo out-of-sample forecast errors is an estimator of the
As discussed in Section 12.4, this estimator of the RMSFE can be used
tify forecast uncertainty and to construct forecast intervals.

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

Application: Did the Phillips curve change during the 1990s
coefficients of the Phillips curve changed during the 1990s, then pse
of-sample forecasts computed over that period should deteriorate. Th
out-of-sample forecasts of inflation for the period 1994:I to 1999:IV, c
using the four-lag Phillips curve, are plotted in Figure 12.6 along with t
values of inflation. For example, the forecast of inflation for 1994:I v
puted by regressing ΔInf_t on $\Delta Inf_{t-1}, \ldots, \Delta Inf_{t-4}$, $Unemp_{t-1}, \ldots,$ $Unem$
an intercept using the data through 1993:IV, then computing the
$\widehat{\Delta Inf}_{t1994:I|1993:IV}$ using these estimated coefficients and the data through
The inflation forecast for 1994:I is then $\widehat{Inf}_{1994:I|1993:IV} = Inf_{1993:IV} + \widehat{\Delta Inf}_{t199}$
This entire procedure was repeated using data through 1994:I to com
forecast $\widehat{Inf}_{1994:II|1994:I}$. Doing this for all 24 quarters from 1994:I–
creates 24 pseudo out-of-sample forecasts, which are plotted in Figu
The pseudo out-of-sample forecast errors are the differences betwe
inflation and its pseudo out-of-sample forecast, that is, the differences bet
two lines in Figure 12.6. For example, in the third quarter of 1997, the
rate rose by 0.9 percentage points, but the pseudo out-of-sample fo
$\Delta Inf_{1997:III}$ was 1.9 percentage points, so the pseudo out-of-sample forecast
$\Delta Inf_{1997:III} - \widehat{\Delta Inf}_{1997:III|1997:II} = 0.9 - 1.9 = -1.0$ percentage points. In oth
a forecaster using the ADL(4,4) model of the Phillips curve, estimated
1997:II, would have forecasted that inflation would increase by 1.9 percenta
in 1997:III, whereas in reality it only increased by 0.9 percentage points.

How do the mean and standard deviation of the pseudo out-of-sam
cast errors compare with the in-sample fit of the model? The standard
the regression of the four-lag Phillips curve fit using data through 1993:I
so based on the in-sample fit we would expect the out-of-sample forec
to have mean zero and root mean square forecast error of 1.47. In fact,

FIGURE 12.6 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 (12.17) generally track actual inflation, but on average the forecasts are higher than actual inflation. This upward bias in the forecasts may have been caused by a decline in the natural rate of unemployment, which would appear as a shift in the intercept of the Phillips curve.

age forecast error is –0.37 and the sample RMSFE is 0.75. Thus the RMSFE of the pseudo out-of-sample forecasts is *less* than predicted by the in-sample fit of the regression. However, the average forecast error is negative rather than zero, that is, on average the forecasts predicted larger increases in inflation (and thus higher inflation) than actually occurred. In fact, the t-statistic testing the hypothesis that the mean out-of-sample forecast error is zero is $t = -2.71$, so the hypothesis that the mean is zero is rejected at the 1% significance level. This suggests that the forecasts were biased over this period, systematically forecasting higher inflation than actually occurred. The finding that the pseudo out-of-sample forecasts are biased is reflected in Figure 12.6: forecasted inflation typically exceeds actual inflation so the average forecast error is negative.

These biased forecasts suggest that the Phillips curve regression was unstable towards the end of this sample, and that this instability led to forecasts of the

change in inflation that were systematically too high. Before using this moc real-time forecasting, it would be important to try to identify the source shift and to incorporate it into a modified version of the Phillips curve mo

Taken together, this bias in the pseudo out-of-sample forecasts and the tion of stability by the QLR statistic (at the 10% level) suggest that the fo Phillips curve has been unstable. This instability was a matter of considerable est during the 1990s and early 2000s because economic forecasters recog that, as seen in Figure 12.6, inflation forecasts based on the Phillips curve too high. Some macroeconomists think that the source of this instability decline in the natural rate of unemployment during the 1990s, which would late into a negative shift in the intercept in the regressions examined here. macroeconomists think that this breakdown is more complete, however, an the entire concept of the Phillips curve—a link between the pressures of demand and overall price inflation—is just an antiquated feature of the information age economy. If you are interested in reading more on this d see the symposium on the Phillips curve in the Winter 1997 issue of the *J of Economic Perspectives.*

Avoiding the Problems Caused by Breaks

The best way to adjust for a break in the population regression function de on the source of that break. If a distinct break occurs at a specific date, this will be detected with high probability by the QLR statistic, and the break da be estimated. Thus the regression function can be estimated using a binary va indicating the two subsamples associated with this break, interacted with the regressors as needed. If all the coefficients break, then this regression takes the of Equation (12.35), where τ is replaced by the estimated break date, $\hat{\tau}$, while i some of the coefficients break, then only the relevant interaction terms app the regression. If there is in fact a distinct break, then inference on the regr coefficients can proceed as usual, for example using the usual normal critical for hypothesis tests based on t-statistics. In addition, forecasts can be produced 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 parameters, the remedy is more difficult, and goes beyond the scope of this b

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

12.8 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 12.5, 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 or not 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 12.6. 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 12.7.

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 e
dynamic causal effects, and that is the topic of the next chapter.

Summary

1. Regression models used for forecasting need not have a causal interpretati
2. A time series variable generally is correlated with one or more of its lagg ues; that is, it is serially correlated.
3. An autoregression of order p is a linear multiple regression model in whi regressors are the first p lags of the dependent variable. The coefficient AR(p) can be estimated by OLS, and the estimated regression function used for forecasting. The lag order p can be estimated using an informati terion such as the BIC.
4. Adding other variables and their lags to an autoregression can improve fo ing performance. Under the least squares assumptions for time series reg (Key Concept 12.6), the OLS estimators have normal distributions in larg ples and statistical inference proceeds the same way as for cross-sectional d
5. Forecast intervals are one way to quantify forecast uncertainty. If the err normally distributed, an approximate 68% forecast interval can be constru the forecast ± an estimate of the root mean squared forecast error.
6. A series that contains a stochastic trend is nonstationary, violating the s least squares assumption in Key Concept 12.6. The OLS estimat t-statistic for the coefficient of a regressor with a stochastic trend can nonstandard distribution, potentially leading to biased estimators, inef forecasts, and misleading inferences. The ADF statistic can be used to t a stochastic trend. A random walk stochastic trend can be eliminated by first differences of the series.
7. If the population regression function changes over time, then OLS est neglecting this instability are unreliable for statistical inference or forecastin QLR statistic can be used to test for a break and, if a discrete break is fou 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 towa end of the sample, to estimate the root mean squared forecast error, and to pare different forecasting models.

Key Terms

first lag (432)
j^{th} lag (432)
first difference (432)
autocorrelation (434)
serial correlation (434)
autocorrelation coefficient (434)
j^{th} autocovariance (435)
autoregression (438)
forecast error (439)
root mean squared forecast error (440)
AR(p) (441)
autoregressive distributed lag model (445)
ADL(p,q) (445)
stationarity (446)
weak dependence (448)
Granger causality test (449)
forecast interval (451)
BIC (453)
AIC (455)
trend (457)
deterministic trend (457)
stochastic trend (457)
random walk (458)
random walk with drift (459)
unit root (460)
spurious regression (461)
Dickey-Fuller statistic (463)
augmented Dickey-Fuller (ADF) statistic (463)
break date (468)
Quandt likelihood ratio (QLR) statistic (469)
pseudo out-of-sample forecast (473)

Review the Concepts

12.1 Look at the plot of the logarithm of real GDP for Japan in Figure 12.2c. 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.

12.2 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 predicts better than the random walk model. Explain how you would examine the analyst's claim that his model is superior.

12.3 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.

12.4 Suppose that you suspected that the intercept in Equation (12.17) 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

***12.1** Suppose that Y_t follows the stationary AR(1) model $Y_t = 2.5 + 0.7Y_t$
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.

b. Compute the first two autocovariances of Y_t.

c. Compute the first two autocorrelations of Y_t.

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

12.2 The index of industrial production (IP_t) is a monthly time series th
sures the quantity of industrial commodities produced in a given
This problem uses data on this index for the United States. All reg
are estimated over the sample period 1960:1 to 2000:12 (that is, J
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 chan
IP, measured in percentage points per annum. Is this correct? V

b. Suppose a forecaster estimates the following AR(4) model for Y

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

Use this AR(4) to forecast the value of Y_t in January 2001 usin
following values of IP for August 2000 through December 200

Date	2000:7	2000:8	2000:9	2000:10	2000:11	20
IP	147.595	148.650	148.973	148.660	148.206	1

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

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

e. Worried that she might have included too few or too many lag
model, the forecaster estimates AR(p) models for $p = 1, \ldots,$ 6

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	29175	28538	28393	28391	28378	28317

***12.3** Using the same data as in Exercise 12.2, 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 12.2? Explain.

12.4 The forecaster in Exercise 12.2 augments her AR(4) model for *IP* growth to include 4 lagged values of ΔR_t where R_t is the interest rate on 3-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.

12.5 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 wh
$Y_{t|t-1}$, where $Y_{t|t-1} = E(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$. (*Hint:* Extend the r
part (a) to conditional expectations.)

c. Show that the errors u_t of an AR(p) (Equation (12.14) in Key
cept 12.3) are serially uncorrelated. (*Hint:* Use Equation (2.2

12.6 In this exercise you will conduct a Monte Carlo experiment that s
phenomenon of spurious regression discussed in Section 12.6. In
Carlo study, artificial data are generated using a computer, then thes
data are used to calculate the statistics being studied. This makes it p
compute the distribution of statistics for known models when mat
expressions for those distributions are complicated (as they are her
unknown. In this exercise, you will generate data so that two series,
are independently distributed random walks. The specific steps are

i. Use your computer to generate a sequence of $T = 100$ i.i.d.
normal random variables. Call these variables $e_1, e_2, \ldots, e_{100}$
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, $a_1, a_2, \ldots, a_{10}$
100 i.i.d. standard normal random variables. Set $X_1 = a_1$ and
$+ a_t$ for $t = 2, 3, \ldots, 100$.

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

Use this algorithm to answer the following questions:

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

b. Repeat (a) 1,000 times, saving each value of R^2 and the t-stat
Construct a histogram of the R^2 and t-statistic. What are the
and 95% percentiles of the distributions of the R^2 and the t-s
In what fraction of your 1,000 simulated data sets does the t-
exceed 1.96 in absolute value?

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

APPENDIX 12.1 Time Series Data Used in Chapter 12

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 real 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 websites maintained by various data collecting agencies.

APPENDIX 12.2 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 12.5 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. 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 (12.8) 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

$$Y_t = u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \beta_1^3 u_{t-3} + \ldots = \sum_{i=0}^{\infty} \beta_1^i u_{t-i}. \quad (12.37)$$

Thus Y_t is a weighted average of current and past u_t's. Because the u_t's are n distributed and because the weighted average of normal random variables is norm tion 2.6), Y_{s+1} and Y_{s+2} have a bivariate normal distribution. Recall from Section the bivariate normal distribution is completely determined by the means of the tw ables, their variances, and their covariance. Thus, to show that Y_t is stationary, we show that the means, variances, and covariance of (Y_{s+1}, Y_{s+2}) do not depend o extension of the argument used below can be used to show that the distribution c $Y_{s+2}, \ldots, Y_{s+T})$ does not depend on s.

The means and variances of Y_{s+1} and Y_{s+2} can be computed using Equation (with the subscripts $s + 1$ or $s + 2$ replacing t. First, because $E(u_t) = 0$ for all t, $E(\sum_{i=0}^{\infty}\beta_1^i u_{t-i}) = \sum_{i=0}^{\infty}\beta_1^i E(u_{t-i}) = 0$, so the mean of Y_{s+1} and Y_{s+2} are both zero and in ular do not depend on s. Second, $\text{var}(Y_t) = \text{var}(\sum_{i=0}^{\infty}\beta_1^i u_{t-i}) = \sum_{i=0}^{\infty}(\beta_1^i)^2\text{var}(u_{t-i}) = \sigma_u^2\sum$ $\sigma_u^2/(1-\beta_1^2)$, where the final equality follows from the fact that, if $|a| < 1, \sum_{i=0}^{\infty} a^i = 1/(1-$ $\text{var}(Y_{s+1}) = \text{var}(Y_{s+2}) = \sigma_u^2/(1-\beta_1^2)$, which does not depend on s as long as $|\beta_1| < 1$. because $Y_{s+2} = \beta_1 Y_{s+1} + u_{s+2}$, $\text{cov}(Y_{s+1}, Y_{s+2}) = E(Y_{s+1}Y_{s+2}) = E[Y_{s+1}(\beta_1 Y_{s+1} +$ $\beta_1\text{var}(Y_{s+1}) + \text{cov}(Y_{s+1}, u_{s+2}) = \beta_1\text{var}(Y_{s+1}) = \beta_1\sigma_u^2/(1-\beta_1^2)$. The covariance does not on s, so Y_{s+1} and Y_{s+2} have a joint probability distribution that does not depend o is, their joint distribution is stationary. If $|\beta_1| \geq 1$, this calculation breaks down beca infinite sum in Equation (12.37) does not converge and the variance of Y_t is infinit Y_t is stationary if $|\beta_1| < 1$, but not if $\beta_1 = 1$.

The preceding argument was made under the assumptions that $\beta_0 = 0$ and u_t mally distributed. If $\beta_0 \neq 0$, the argument is similar except that the means of Y_{s+1} a are $\beta_0/(1-\beta_1)$ and Equation (12.37) must be modified for this nonzero mea assumption that u_t is i.i.d. normal can be replaced with the assumption that u_t is ary with a finite variance because, by Equation (12.37), Y_t can still be expressed as tion of current and past u_t's, so the distribution of Y_t is stationary as long as the distr of u_t is stationary and the infinite sum expression in Equation (12.37) is meaningfu sense that it converges, which requires $|\beta_1| < 1$.

APPENDIX 12.3 Lag Operator Notation

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

$$LY_t = Y_{t-1},\ L^2Y_t = Y_{t-2},\ \text{and}\ L^jY_t = Y_{t-j}. \tag{12.38}$$

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

$$a(L) = a_0 + a_1L + a_2L^2 + \cdots + a_pL^p = \sum_{j=0}^{p} a_jL^j, \tag{12.39}$$

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 (12.39) is p. Multiplying Y_t by $a(L)$ yields

$$a(L)Y_t = \left(\sum_{j=0}^{p} a_jL^j\right)Y_t = \sum_{j=0}^{p} a_j(L^jY_t) = \sum_{j=0}^{p} a_jY_{t-j} = a_0Y_t + a_1Y_{t-1} + \cdots + a_pY_{t-p}. \tag{12.40}$$

The expression in Equation (12.40) implies that the AR(p) model in Equation (12.14) can be written compactly as

$$a(L)Y_t = \beta_0 + u_t, \tag{12.41}$$

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

$$a(L)Y_t = \beta_0 + c(L)X_{t-1} + u_t, \tag{12.42}$$

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 12.4 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. That is, in the lag operator notation of

Appendix 12.3, let $u_t = b(L)e_t$, where e_t is a serially uncorrelated, unobserved able, and $b(L)$ is a lag polynomial of degree q with $b_0 = 1$. Then the ARMA

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

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 approxima variances of Y_t. The reason for this is that any stationary time series Y_t wit ance can be written either as an AR or as a MA with a serially uncorrelat although the AR or MA models might need to have an infinite order. The s results, that a stationary process can be written in moving average form, is Wold decomposition theorem, and is one of the fundamental results unde ory of stationary time series analysis.

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

APPENDIX 12.5 Consistency of the BIC Lag Length Esti

This appendix summarizes the argument that the BIC estimator of the la an autoregression is correct in large samples, that is, $\Pr(\hat{p} = p) \longrightarrow 1$. This 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 a with zero, one, or two lags, when the true lag length is one. It is shov (i) $\Pr(\hat{p} = 0) \longrightarrow 0$, and (ii) $\Pr(\hat{p} = 2) \longrightarrow 0$, from which it follows that P The extension of this argument to the general case of searching over $0 \leq$ showing that $\Pr(\hat{p} < p) \longrightarrow 0$ and $\Pr(\hat{p} > p) \longrightarrow 0$; the strategy for showi 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 $BIC(0) - BIC(1) = [\ln(SSR(0)/T) + (\ln T)/T] - [\ln(SSR(1)/T) + 2(\ln T)/T] = \ln(SSR(0)/T) - \ln(SSR(1)/T) - (\ln T)/T$. Now $SSR(0)/T = [(T-1)/T]s_Y^2 \xrightarrow{p} \sigma_Y^2$, $SSR(1)/T \xrightarrow{p} \sigma_u^2$, and $(\ln T)/T \longrightarrow 0$; putting these pieces together, $BIC(0) - BIC(1) \xrightarrow{p} \ln\sigma_Y^2 - \ln\sigma_u^2 > 0$ because $\sigma_Y^2 > \sigma_u^2$. It follows that $\Pr[BIC(0) < BIC(1)] \longrightarrow 0$, so that $\Pr(\hat{p} = 0) \longrightarrow 0$.

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 $T[BIC(2) - BIC(1)] = T\{[\ln(SSR(2)/T) + 3(\ln T)/T] - [\ln(SSR(1)/T) + 2(\ln T)/T]\} = T\ln[SSR(2)/SSR(1)] + \ln T = -T\ln[1 + F/(T-2)] + \ln T$, where $F = [SSR(1) - SSR(2)]/[SSR(2)/(T-2)]$ is the "rule of thumb" F-statistic (Appendix 5.3) testing the null hypothesis that $\beta_2 = 0$ in the AR(2). If u_t is homoskedastic, F has a χ_1^2 asymptotic distribution; if not, it has some other asymptotic distribution. Thus $\Pr[BIC(2) - BIC(1) < 0] = \Pr[T(BIC(2) - BIC(1)) < 0] = \Pr\{-T\ln[1 + F/(T-2)] + (\ln T) < 0\} = \Pr\{T\ln[1 + F/(T-2)] > \ln T\}$. As T increases, $T\ln[1 + F/(T-2)] - F \longrightarrow 0$ (a consequence of the logarithmic approximation $\ln(1+a) \cong a$, which becomes exact as $a \longrightarrow 0$). Thus $\Pr[BIC(2) - BIC(1) < 0] \longrightarrow \Pr(F > \ln T) \longrightarrow 0$, so that $\Pr(\hat{p} = 2) \longrightarrow 0$.

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 $\Pr(AIC(2) - AIC(1) < 0) \longrightarrow \Pr(F > 2) > 0$. If u_t is homoskedastic, $\Pr(F > 2) \longrightarrow \Pr(\chi_1^2 > 2) = 0.16$, so that $\Pr(\hat{p} = 2) \longrightarrow 0.16$. 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 13

Estimation of Dynamic Causal Effects

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

This chapter takes up the problem of estimating the effect on Y now
the future of a change in X, that is, the **dynamic causal effect** on Y of
in X. What, for example, is the effect on the path of orange juice prices
of a freezing spell in Florida? The starting point for modeling and estima
dynamic causal effects is the so-called distributed lag regression model, i
Y_t is expressed as a function of current and past values of X_t. Section 13.
introduces the distributed lag model in the context of estimating the eff
weather in Florida on the price of orange juice concentrate over time. S
13.2 takes a closer look at what, precisely, is meant by a dynamic causal

One way to estimate dynamic causal effects is to estimate the coeff
the distributed lag regression model using OLS. As discussed in Sectio
this estimator is consistent if the regression error has a conditional mea
given current and past values of X, a condition that (as in Chapter 10)

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 new, "heteroskedasticity- and autocorrelation-consistent" (HAC) formulas for standard errors, the topic of Section 13.4.

A second way to estimate dynamic causal effects, discussed in Section 13.5, 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 by generalized least squares (GLS). Both the ADL and GLS methods, however, require a stronger version of exogeneity than we have used so far: *strict* exogeneity, under which the regression errors have a conditional mean of zero given past, present, *and future* values of X.

Section 13.6 provides a more complete analysis of the relationship between orange juice prices and the weather. In this application, the weather is beyond human control and thus is exogenous (although, as discussed in Section 13.6, economic theory suggests that it is not necessarily strictly exogenous). Because exogeneity is necessary for estimating dynamic causal effects, Section 13.7 examines this assumption in several applications taken from macroeconomics and finance.

This chapter builds on the material in Sections 12.1–12.4 but, with the exception of a subsection (that can be skipped) of the empirical analysis in Section 13.6, does not require the material in Sections 12.5–12.8.

13.1 An Initial Taste of the Orange Juice Data

Orlando, the center of Florida's orange growing region, is normally sunny and warm. But now and then there is a cold snap, and if temperatures drop below freezing for too long the trees drop many of their oranges and, if the freeze is severe, the trees freeze. Following a freeze, the supply of orange juice concentrate

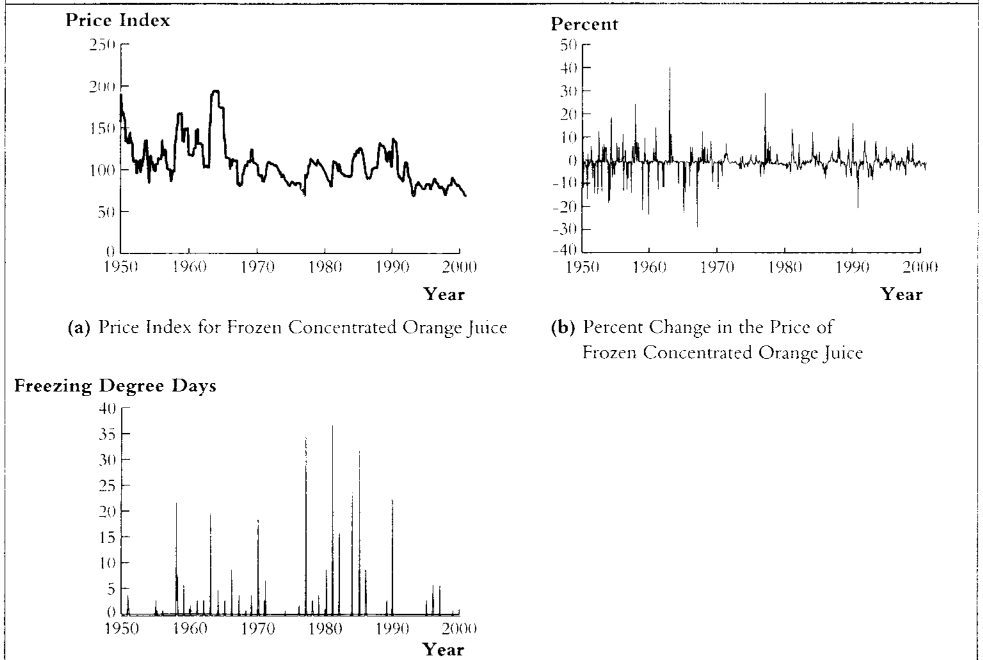
falls and its price rises. The timing of the price increases is rather com[p]
however. Orange juice concentrate is a "durable," or storable, commodity
it can be stored in its frozen state, albeit at some cost (to run the freeze[r]
the price of orange juice concentrate depends not only on current supply
on expectations of future supply. A freeze today means that future supplies
centrate will be low, but because concentrate currently in storage can be
meet either current or future demand, the price of existing concentrate ris[es]
But precisely how much does the price of concentrate rise when there is [a]
The answer to this question is of interest not just to orange juice traders b[ut]
generally to economists interested in studying the operations of moder[n]
modity markets. To learn how the price of orange juice changes in resp[onse]
weather conditions, we must analyze data on orange juice prices and the [weather].

Monthly data on the price of frozen orange juice concentrate, its [monthly]
percentage change, and temperatures in the orange growing region of [Florida]
from January 1950 to December 2000 are plotted in Figure 13.1. The pri[ce]
ted in Figure 13.1a, is a measure of the average real price of frozen oran[ge]
concentrate paid by wholesalers. This price was deflated by the overall p[roducer]
price index for finished goods to eliminate the effects of overall price i[nflation].
The percentage price change plotted in Figure 13.1b is the change in t[he price]
over the month. The temperature data plotted in Figure 13.1c are the nu[mber of]
"freezing degree days" at the Orlando, Florida, airport, calculated as the
the number of degrees Fahrenheit that the minimum temperature fall[s below]
freezing in a given day over all days in the month; for example, in Novemb[er]
the airport temperature dropped below freezing twice, on the 25th (31°)
the 29th (29°) for a total of four freezing degree days ((32 − 31) + (32 − 2[9])
(The data are described in more detail in Appendix 13.1.) As you can see
paring the panels in Figure 13.1, the price of orange juice concentrate
swings, some of which appear to be associated with cold weather in Flo[rida].

We begin our quantitative analysis of the relationship between oran[ge juice]
price and the weather by using a regression to estimate the amount b[y which]
orange juice prices rise when the weather turns cold. The dependent va[riable is]
the percentage change in the price over that month ($\%ChgP_t$, where $\%$
$100 \times \Delta\ln(P_t^{OJ})$ and P_t^{OJ} is the real price of orange juice). The regressor is t[he num-]
ber of freezing degree days during that month (FDD_t). This regressio[n is esti-]
mated using monthly data from January 1950 to December 2000 (a[s are all]
regressions in this chapter), for a total of $T = 612$ observations:

$$\widehat{\%ChgP_t} = -0.40 + 0.47FDD_t.$$
$$(0.22)\quad(0.13)$$

FIGURE 13.1 Orange Juice Prices and Florida Weather, 1950–2000



(a) Price Index for Frozen Concentrated Orange Juice

(b) Percent Change in the Price of Frozen Concentrated Orange Juice

(c) Monthly Freezing Degree Days in Orlando, Florida

There have been large month-to-month changes in the price of frozen concentrated orange juice. Many of the large movements coincide with freezing weather in Orlando, home of the orange groves.

The standard errors reported in this section are not the usual OLS standard errors, but rather are heteroskedasticity- and autocorrelation-consistent (HAC) standard errors that are appropriate when the error term and regressors are autocorrelated. HAC standard errors are discussed in Section 13.4, and for now they are used without further explanation.

According to this regression, an additional freezing degree day during a month increases the price of orange juice concentrate over that month by 0.47%. In a month with four freezing degree days, such as November 1950, the price of orange juice concentrate is estimated to have increased by 1.88% (4 × 0.47% = 1.88%), relative to a month with no days below freezing.

Because the regression in Equation (13.1) includes only a contempo measure of the weather, it does not capture any lingering effects of the c on the orange juice price over the coming months. To capture these we consider the effect on prices of both contemporaneous and lagged values which in turn can be done by augmenting the regression in Equation (13 for example, lagged values of FDD over the previous six months:

$$\widehat{\%ChgP_t} = \underset{(0.23)}{-0.65} + \underset{(0.14)}{0.47}FDD_t + \underset{(0.08)}{0.14}FDD_{t-1} + \underset{(0.06)}{0.06}FDD_{t-2}$$
$$+ \underset{(0.05)}{0.07}FDD_{t-3} + \underset{(0.03)}{0.03}FDD_{t-4} + \underset{(0.03)}{0.05}FDD_{t-5} + \underset{(0.04)}{0.05}FDD_{t-6}.$$

Equation (13.2) is a distributed lag regression. The coefficient on FDD_t tion (13.2) estimates the percentage increase in prices over the course of th in which the freeze occurs; an additional freezing degree day is estimated to prices that month by 0.47%. The coefficient on the first lag of FDD_t, FDI mates the percentage increase in prices arising from a freezing degree day in ceding month, the coefficient on the second lag estimates the effect of a degree day two months ago, and so forth. Equivalently, the coefficient on lag of FDD estimates the effect of a unit increase in FDD one month after t occurs. Thus the estimated coefficients in Equation (13.2) are estimates of t of a unit increase in FDD_t on current and future values of $\%ChgP$, that is, estimates of the dynamic effect of FDD_t on $\%ChgP_t$. For example, the fo ing degree days in November 1950 are estimated to have increased oran prices by 1.88% during November 1950, by an additional 0.56% (= 4 × December 1950, by an additional 0.24% (= 4 × 0.06) in January 1951, and

13.2 Dynamic Causal Effects

Before learning more about the tools for estimating dynamic causal eff should spend a moment thinking about what, precisely, is meant by a causal effect. Having a clear idea about what a dynamic causal effect is l clearer understanding of the conditions under which it can be estimated

Causal Effects and Time Series Data

Section 1.2 defined a causal effect as the outcome of an ideal randomiz trolled experiment: when a horticulturalist randomly applies fertilizer tomato plots but not others and then measures the yield, the expected difference in yield between the fertilized and unfertilized plots is the effect on tomato yield of the fertilizer. This concept of an experiment, however, is one in which there are multiple subjects (multiple tomato plots or multiple people), so the data are either cross-sectional (the tomato yield at the end of the harvest) or panel data (individual incomes before and after an experimental job training program). By having multiple subjects, it is possible to have both treatment and control groups and thereby to estimate the causal effect of the treatment.

In time series applications, this definition of causal effects in terms of an ideal randomized controlled experiment needs to be modified. To be concrete, consider an important problem of macroeconomics: estimating the effect of an unanticipated change in the short-term interest rate on the current and future economic activity in a given country, as measured by GDP. Taken literally, the randomized controlled experiment of Section 1.2 would entail randomly assigning different economies to treatment and control groups. The central banks in the treatment group would apply the treatment of a random interest rate change, while those in the control group would apply no such random changes; for both groups, economic activity (for example, GDP) would be measured over the next few years. But what if we are interested in estimating this effect for a specific country, say the United States? Then this experiment would entail having different "copies" of the United States as subjects, and assigning some copies to the treatment and some to the control group. Obviously, this "parallel universes" experiment is infeasible.

Instead, in time series data it is useful to think of a randomized controlled experiment consisting of the same subject (e.g., the U.S. economy) being given different treatments (randomly chosen changes in interest rates) at different points in time (the 1970s, the 1980s, and so forth). In this framework, the single subject at different times plays the role of both treatment and control group: sometimes the Fed changes the interest rate while at other times it does not. Because data are collected over time, it is possible to measure the dynamic causal effect, that is, the time path of the effect on the outcome of interest of the treatment. For example, a surprise increase in the short-term interest rate of two percentage points, sustained for one quarter, might initially have a negligible effect on output; after two quarters GDP growth might slow, with the greatest slowdown after one and one-half years; then over the next two years, GDP growth might return to normal. This time path of causal effects is the dynamic causal effect on GDP growth of a surprise change in the interest rate.

As a second example, consider the causal effect on orange juice price changes of a freezing degree day. It is possible to imagine a variety of hypothetical experiments, each yielding a different causal effect. One experiment would be to

change the weather in the Florida orange groves, holding constant weathe where—for example, holding constant weather in the Texas grapefruit grov in other citrus fruit regions. This experiment would measure a partial effect, ing other weather constant. A second experiment might change the weat all the regions, where the "treatment" is application of overall weather pa If weather is correlated across regions for competing crops, then thes dynamic causal effects differ. In this chapter, we consider the causal effect latter experiment, that is, the causal effect of applying general weather pa This corresponds to measuring the dynamic effect on prices of a change in F weather, *not* holding constant weather in other agricultural regions.

Dynamic effects and the distributed lag model. Because dynamic necessarily occur over time, the econometric model used to estimate dy causal effects needs to incorporate lags. To do so, Y_t can be expressed as a d uted lag of current and r past values of X_t:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t,$$

where u_t is an error term that includes measurement error in Y_t and the ef omitted determinants of Y_t. The model in Equation (13.3) is called the **di uted lag model** relating X_t, and r of its lags, to Y_t.

As an illustration of Equation (13.3), consider a modified version tomato/fertilizer experiment: because fertilizer applied today might remain ground in future years, the horticulturalist wants to determine the effect on t yield *over time* of applying fertilizer. Accordingly, she designs a three-year e ment and randomly divides her plots into four groups: the first is fertilized i the first year; the second is fertilized in only the second year; the third is fer in only the third year; and the fourth, the control group, is never fertilized. toes are grown annually in each plot, and the third-year harvest is weighed three treatment groups are denoted by the binary variables X_{t-2}, X_{t-1}, a where t represents the third year (the year in which the harvest is weighed $= 1$ if the plot is in the first group (fertilized two years earlier), $X_{t-1} = 1$ if th was fertilized one year earlier, and $X_t = 1$ if the plot was fertilized in the fina In the context of Equation (13.3) (which applies to a single plot), the effect of fertilized in the final year is β_1, the effect of being fertilized one year earlie and the effect of being fertilized two years earlier is β_3. If the effect of fertil greatest in the year it is applied, then β_1 would be larger than β_2 and β_3.

More generally, the coefficient on the contemporaneous value of X_t, β_1, contemporaneous or immediate effect of a unit change in X_t on Y_t. The coef on X_{t-1}, β_2, is the effect on Y_t of a unit change in X_{t-1} or, equivalently, the effect on Y_{t+1} of a unit change in X_t; that is, β_2 is the effect of a unit change in X on Y one period later. In general, the coefficient on X_{t-h} is the effect of a unit change in X on Y after h periods. The dynamic causal effect is the effect of a change in X_t on Y_t, Y_{t+1}, Y_{t+2}, and so forth, that is, it is the sequence of causal effects on current and future values of Y. Thus, in the context of the distributed lag model in Equation (13.3), the dynamic causal effect is the sequence of coefficients $\beta_1, \beta_2, \ldots, \beta_{r+1}$.

Implications for empirical time series analysis. This formulation of dynamic causal effects in time series data as the expected outcome of an experiment in which different treatment levels are repeatedly applied to the same subject has two implications for empirical attempts to measure the dynamic causal effect with observational time series data. The first implication is that the dynamic causal effect should not change over the sample on which we have data. This in turn is implied by the data being jointly stationary (Key Concept 12.5). As discussed in Section 12.7, the hypothesis that a population regression function is stable over time can be tested using the QLR test for a break, in which case it is possible to estimate the dynamic causal effect in different subsamples. The second implication is that X must be uncorrelated with the error term, and it is to this implication that we now turn.

Two Types of Exogeneity

Section 10.1 defined an "exogenous" variable to be a variable that is uncorrelated with the regression error term and an "endogenous" variable to be a variable that is correlated with the error term. This terminology traces to models with multiple equations, in which an "endogenous" variable is determined within the model while an "exogenous" variable is determined outside the model. Loosely speaking, if we are to estimate dynamic causal effects using the distributed lag model in Equation (13.3), the regressors (the X's) must be uncorrelated with the error term. Thus, X must be exogenous. Because we are working with time series data, however, we need to refine the definitions of exogeneity. In fact, there are two different concepts of exogeneity that we use here.

The first concept of exogeneity is that the error term has a conditional mean of zero given current and all past values of X_t, that is, that $E(u_t | X_t, X_{t-1}, X_{t-2}, \ldots) = 0$. This modifies the standard conditional mean assumption for multiple regression with cross-sectional data (Assumption 1 in Key Concept 5.4), which requires only that u_t has a conditional mean of zero given the included regressors; that is, that $E(u_t | X_t, X_{t-1}, \ldots, X_{t-r}) = 0$. Including all lagged values of X_t in the conditional

expectation implies that all the more distant causal effects—all the causal
beyond lag r—are zero. Thus, under this assumption, the r distributed lag c
cients in Equation (13.3) constitute all of the nonzero dynamic causal effect
can refer to this assumption—that $E(u_t | X_t, X_{t-1}, \ldots) = 0$—as **past and pr**
exogeneity, but because of the similarity of this definition and the definiti
exogeneity in Chapter 10, we just use the term **exogeneity**.

The second concept of exogeneity is that the error term has mean zero,
all past, present, and *future* values of X_t, that is, that $E(u_t | \ldots, X_{t+2}, X_{t+1}, X_t$
$X_{t-2}, \ldots) = 0$. This is called **strict exogeneity**; for clarity, we also call it
present, and future exogeneity. The reason for introducing the conce
strict exogeneity is that, when X is strictly exogenous, there are more eff
estimators of dynamic causal effects than the OLS estimators of the coeffi
of the distributed lag regression in Equation (13.3).

The difference between exogeneity (past and present) and strict exoge
(past, present, and future) is that strict exogeneity includes future values of
the conditional expectation. Thus, strict exogeneity implies exogeneity, bu
vice versa. One way to understand the difference between the two conce
to consider the implications of these definitions for correlations between X
u. If X is (past and present) exogenous, then u_t is uncorrelated with curren
past values of X_t. If X is strictly exogenous, then in addition u_t is uncorre
with *future* values of X_t. For example, if a change in Y_t causes *future* values
to change, then X_t is not strictly exogenous even though it might be (pas
present) exogenous.

As an illustration, consider the hypothetical multiyear tomato/fertilizer ex
iment described following Equation (3.3). Because the fertilizer is rando
applied in the hypothetical experiment, it is exogenous. Because tomato
today does not depend on the amount of fertilizer applied in the future, the
tilizer time series is also strictly exogenous.

As a second illustration, consider the orange juice price example, in whi
is the monthly percentage change in orange juice prices and X_t is the numb
freezing degree days in that month. From the perspective of orange juice mar
we can think of the weather—the number of freezing degree days—as if it
randomly assigned, in the sense that the weather is outside human control. I
effect of FDD is linear and if it has no effect on prices after r months, then it
lows that the weather is exogenous. But is the weather *strictly* exogenous? I
conditional mean of u_t given future FDD is nonzero, then FDD is not str
exogenous. To answer this question requires thinking carefully about w
precisely, is contained in u_t. In particular, if OJ market participants use fore
of FDD when they decide how much they will buy or sell at a given price,

Key Concept 13.1

The Distributed Lag Model and Exogeneity

In the distributed lag model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (13.4)$$

there are two different types of exogeneity, that is, two different exogeneity conditions:

Past and present exogeneity (exogeneity):

$$E(u_t | X_t, X_{t-1}, X_{t-2}, \ldots) = 0; \quad (13.5)$$

Past, present, and future exogeneity (strict exogeneity):

$$E(u_t | \ldots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0. \quad (13.6)$$

If X is strictly exogenous it is exogenous, but exogeneity does not imply strict exogeneity.

OJ prices, and thus the error term u_t, could incorporate information about future *FDD* that would make u_t a useful predictor of *FDD*. This means that u_t will be correlated with future values of FDD_t. According to this logic, because u_t includes forecasts of future Florida weather, *FDD* would be (past and present) exogenous but not *strictly* exogenous. The difference between this and the tomato/fertilizer example is that, while tomato plants are unaffected by future fertilizing, OJ market participants *are* influenced by forecasts of future Florida weather. We return to the question of whether *FDD* is strictly exogenous when we analyze the orange juice price data in more detail in Section 13.6.

The two definitions of exogeneity are summarized in Key Concept 13.1.

13.3 Estimation of Dynamic Causal Effects with Exogenous Regressors

If X is exogenous, then its dynamic causal effect on Y can be estimated by OLS estimation of the distributed lag regression in Equation (13.4). This section summarizes the conditions under which these OLS estimators lead to valid statistical inferences and introduces dynamic multipliers and cumulative dynamic multipliers.

The Distributed Lag Model Assumptions

The four assumptions of the distributed lag regression model are similar four assumptions for the cross-sectional multiple regression model (Key Co 5.4), modified for time series data.

The first assumption is that X is exogenous, which extends the zero tional mean assumption for cross-sectional data to include all lagged value As discussed in Section 13.2, this assumption implies that the r distributed lag ficients in Equation (13.3) constitute all of the nonzero dynamic causal effe this sense, the population regression function summarizes the entire dynamic on Y of a change in X.

The second assumption has two parts: part (a) requires that the variable a stationary distribution, and part (b) requires that they become indepen distributed when the amount of time separating them becomes large assumption is the same as the corresponding assumption for the ADL mod second assumption in Key Concept 12.6), and the discussion of this assum in Section 12.4 applies here as well.

The third assumption is that the variables have more than eight nonzero, moments. This is stronger than the assumption of four finite moments that i elsewhere in this book. As discussed in Section 13.4, this stronger assump used in the mathematics behind the HAC variance estimator.

The fourth assumption, which is the same as in the cross-sectional mu regression model, is that there is no perfect multicollinearity.

The distributed lag regression model and assumptions are summarized i Concept 13.2.

Extension to additional X's. The distributed lag model extends direc multiple X's: the additional X's and their lags are simply included as regr in the distributed lag regression, and the assumptions in Key Concept 13 modified to include these additional regressors. Although the extension to tiple X's is conceptually straightforward, it complicates the notation, obsc the main ideas of estimation and inference in the distributed lag model. Fo reason, the case of multiple X's is not treated explicitly in this chapter but as a straightforward extension of the distributed lag model with a single

Autocorrelated u_t, Standard Errors, and Inference

In the distributed lag regression model, the error term u_t can be autocorre that is, u_t can be correlated with its lagged values. This autocorrelation

Key Concept 13.2

The Distributed Lag Model Assumptions

The distributed lag model is given in Key Concept 13.1 (Equation (13.4)), where

1. X is exogenous, that is, $E(u_t | X_t, X_{t-1}, X_{t-2}, \ldots) = 0$;
2. (a) The random variables Y_t and X_t have a stationary distribution, and
   (b) (Y_t, X_t) and (Y_{t-j}, X_{t-j}) become independent as j gets large;
3. Y_t and X_t have more than eight nonzero, finite moments; and
4. There is no perfect multicollinearity.

because, in time series data, the omitted factors included in u_t can themselves be serially correlated. For example, suppose that the demand for orange juice also depends on income, so that one factor that influences the price of orange juice is income, specifically, the aggregate income of potential orange juice consumers. Then aggregate income is an omitted variable in the distributed lag regression of orange juice price changes against freezing degree days. Aggregate income, however, is serially correlated: income tends to fall in recessions and rise in expansions. Thus, income is serially correlated, and, because it is part of the error term, u_t will be serially correlated. This example is typical: because omitted determinants of Y are themselves serially correlated, in general u_t in the distributed lag model will be correlated.

The autocorrelation of u_t does not affect the consistency of OLS, nor does it introduce bias. If, however, the errors are autocorrelated, then in general the usual OLS standard errors are inconsistent and a different formula must be used. Thus correlation of the errors is analogous to heteroskedasticity: the homoskedasticity-only standard errors are "wrong" when the errors are in fact heteroskedastic, in the sense that using homoskedasticity-only standard errors results in misleading statistical inferences when the errors are heteroskedastic. Similarly, when the errors are serially correlated, standard errors predicated upon i.i.d. errors are "wrong" in the sense that they result in misleading statistical inferences. The solution to this problem is to use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors, the topic of Section 13.4.

Dynamic Multipliers and Cumulative Dynamic Multipliers

Another name for the dynamic causal effect is the dynamic multiplier. Th[e] lative dynamic multipliers are the cumulative causal effects, up to a given [...] the cumulative dynamic multipliers measure the cumulative effect o[n] change in X.

Dynamic multipliers. The effect of a unit change in X on Y after h [...] which is β_{h+1} in Equation (13.4), is called the h-period **dynamic mul[tiplier]** Thus, the dynamic multipliers relating X to Y are the coefficients on X [...] lags in Equation (13.4). For example, β_2 is the one-period dynamic multi[plier] is the two-period dynamic multiplier, and so forth. In this terminology, t[he] period (or contemporaneous) dynamic multiplier, or **impact effect**, is [...] effect on Y of a change in X in the same period.

Because the dynamic multipliers are estimated by the OLS regressio[n] ficients, their standard errors are the HAC standard errors of the OLS [...] sion coefficients.

Cumulative dynamic multipliers. The h-period **cumulative dy[namic] multiplier** is the cumulative effect of a unit change in X on Y over the [...] periods. Thus, the cumulative dynamic multipliers are the cumulative sum [...] dynamic multipliers. In terms of the coefficients of the distributed lag reg[ression] in Equation (13.4), the zero-period cumulative multiplier is β_1, the one [...] cumulative multiplier is $\beta_1 + \beta_2$, and the h-period cumulative dynamic mu[ltiplier] is $\beta_1 + \beta_2 + \cdots + \beta_{h+1}$. The sum of all the individual dynamic multiplie[rs] $\beta_2 + \cdots + \beta_{r+1}$, is the cumulative long-run effect on Y of a change in X [...] called the **long-run cumulative dynamic multiplier.**

For example, consider the regression in Equation (13.2). The immedia[te] of an additional freezing degree day is that the price of orange juice conc[entrate] rises by 0.47%. The cumulative effect of a price change over the next m[onth] the sum of the impact effect and the dynamic effect one month ahead; t[he] cumulative effect on prices is the initial increase of 0.47% plus the subs[equent] smaller increase of 0.14% for a total of 0.61%. Similarly, the cumulative d[ynamic] multiplier over two months is 0.47% + 0.14% + 0.06% = 0.67%.

The cumulative dynamic multipliers can be estimated directly using a m[odifica]tion of the distributed lag regression in Equation (13.4). This modified regre[ssion]

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \delta_3 \Delta X_{t-2} + \cdots + \delta_r \Delta X_{t-r+1} + \delta_{r+1} X_{t-r} + u_t$$

The coefficients in Equation (13.7), $\delta_1, \delta_2, \ldots, \delta_{r+1}$ are in fact the cumulative dynamic multipliers. This can be shown by a bit of algebra (Exercise 13.5), which demonstrates that the population regressions in Equations (13.7) and (13.4) are equivalent, where $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$, and so forth. The coefficient on X_{t-r}, δ_{r+1}, is the long-run cumulative dynamic multiplier, that is, $\delta_{r+1} = \beta_1 + \beta_2 + \beta_3 + \cdots + \beta_{r+1}$. Moreover, the OLS estimators of the coefficients in Equation (13.7) are the same as the corresponding cumulative sum of the OLS estimators in Equation (13.4). For example, $\hat{\delta}_2 = \hat{\beta}_1 + \hat{\beta}_2$. The main benefit of estimating the cumulative dynamic multipliers using the specification in Equation (13.7) is that, because the OLS estimators of the regression coefficients are estimators of the cumulative dynamic multipliers, the HAC standard errors of the coefficients in Equation (13.7) are the HAC standard errors of the cumulative dynamic multipliers.

13.4 Heteroskedasticity- and Autocorrelation-Consistent Standard Errors

If the error term u_t is autocorrelated, then OLS is consistent, but in general the usual OLS standard errors for cross-sectional data are not. This means that conventional statistical inferences—hypothesis tests and confidence intervals—based on the usual OLS standard errors will, in general, be misleading. For example, confidence intervals constructed as the OLS estimator ±1.96 conventional standard errors need not contain the true value in 95% of repeated samples, even if the sample size is large. This section begins with a derivation of the correct formula for the variance of the OLS estimator with autocorrelated errors, then turns to heteroskedasticity and autocorrelation-consistent standard errors.

Distribution of the OLS Estimator with Autocorrelated Errors

To keep things simple, consider the OLS estimator $\hat{\beta}_1$ in the distributed lag regression model with no lags, that is, the linear regression model with a single regressor X_t:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad (13.8)$$

where the assumptions of Key Concept 13.2 are satisfied. This section shov the variance of $\hat{\beta}_1$ can be written as the product of two terms: the expressi var($\hat{\beta}_1$), applicable if u_t is not serially correlated, times a correction factor tha from the autocorrelation in u_t or, more precisely, the autocorrelation in (X_t –

As shown in Appendix 4.3, the formula for the OLS estimator $\hat{\beta}_1$ in Key cept 4.2 can be rewritten as

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})u_t}{\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2},$$

where Equation (13.9) is Equation (4.51) with a change of notation so that are replaced by t and T. Because $\overline{X} \xrightarrow{p} \mu_X$ and $\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2 \xrightarrow{p} \sigma_X^2$, ir samples $\hat{\beta}_1 - \beta_1$ is approximately given by

$$\hat{\beta}_1 - \beta_1 \cong \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu_X)u_t}{\sigma_X^2} = \frac{\frac{1}{T}\sum_{t=1}^{T}v_t}{\sigma_X^2} = \frac{\overline{v}}{\sigma_X^2}, \quad ($$

where $v_t = (X_t - \mu_X)u_t$ and $\overline{v} = \frac{1}{T}\sum_{t=1}^{T}v_t$. Thus,

$$\text{var}(\hat{\beta}_1) = \text{var}\left(\frac{\overline{v}}{\sigma_X^2}\right) = \frac{\text{var}(\overline{v})}{(\sigma_X^2)^2}. \quad ($$

If v_t is i.i.d.—as assumed for cross-sectional data in Key Concept 4.3– var($\overline{v}$) = var(v_t)/T and the formula for the variance of $\hat{\beta}_1$ from Key Conce applies. If, however, u_t and X_t are not independently distributed over time in general v_t will be serially correlated, so the formula for the variance of $\overline{v}$ Key Concept 4.4 does not apply. Instead, if v_t is serially correlated, the va of $\overline{v}$ is given by

$$\begin{aligned}\text{var}(\overline{v}) &= \text{var}[(v_1 + v_2 + \cdots + v_T)/T] \\ &= [\text{var}(v_1) + \text{cov}(v_1,v_2) + \cdots + \text{cov}(v_1,v_T) \\ &\quad + \text{cov}(v_2,v_1) + \text{var}(v_2) + \cdots + \text{var}(v_T)]/T^2 \quad (\\ &= [T\text{var}(v_t) + 2(T-1)\text{cov}(v_t,v_{t-1}) + 2(T-2)\text{cov}(v_t,v_{t-2}) \\ &\quad + \cdots + 2\text{cov}(v_t,v_{t-T+1})]/T^2 \\ &= \frac{\sigma_v^2}{T}f_T,\end{aligned}$$

where

$$f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j, \tag{13.13}$$

where $\rho_j = \text{corr}(v_t, v_{t-j})$. In large samples, f_T tends to the limit, $f_T \to f_\infty = 1 + 2\sum_{j=1}^{\infty}\rho_j$.

Combining the expressions in Equation (13.10) for $\hat{\beta}_1$ and Equation (13.12) for $\text{var}(\bar{v})$ gives the formula for the variance of $\hat{\beta}_1$ when v_t is autocorrelated:

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{T}\frac{\sigma_v^2}{(\sigma_X^2)^2}\right]f_T, \tag{13.14}$$

where f_T is given in Equation (13.13).

Equation (13.14) expresses the variance of $\hat{\beta}_1$ as the product of two terms. The first, in square brackets, is the formula for the variance of $\hat{\beta}_1$ given in Key Concept 4.4, which applies in the absence of serial correlation. The second is the factor f_T, which adjusts this formula for serial correlation. Because of this additional factor f_T in Equation (13.14), the OLS standard errors computed using the formula in Key Concept 4.4 are incorrect if the errors are serially correlated: more precisely, if $v_t = (X_t - \mu_X)u_t$ is serially correlated, the estimator of the variance is off by the factor f_T.

HAC Standard Errors

If the factor f_T, defined in Equation (13.13), was known, then the variance of $\hat{\beta}_1$ could be estimated by multiplying the usual cross-sectional estimator of the variance by f_T. This factor, however, depends on the unknown autocorrelations of v_t, so it must be estimated. The estimator of the variance of $\hat{\beta}_1$ that incorporates this adjustment is consistent whether or not there is heteroskedasticity and whether or not v_t is autocorrelated. Accordingly, this estimator is called the **heteroskedasticity- and autocorrelation-consistent (HAC)** estimator of the variance of $\hat{\beta}_1$, and the square root of the HAC variance estimator is the **HAC standard error** of $\hat{\beta}_1$.

The HAC variance formula. The heteroskedasticity- and autocorrelation-consistent estimator of the variance of $\hat{\beta}_1$ is

$$\widetilde{\sigma}^2_{\hat{\beta}_1} = \hat{\sigma}^2_{\hat{\beta}_1}\hat{f}_T, \tag{13.15}$$

where $\hat{\sigma}^2_{\hat{\beta}_1}$ is the estimator of the variance of $\hat{\beta}_1$ in the absence of serial corre tion, given in Equation (4.19), and where $\hat{f}_T$ is an estimator of the factor f_T Equation (13.13).

The task of constructing a consistent estimator $\hat{f}_T$ is challenging. To see w consider two extremes. At one extreme, given the formula in Equation (13.1 it might seem natural to replace the population autocorrelations ρ_j with the sa ple autocorrelations $\hat{\rho}_j$ (defined in Equation (12.6)), yielding the estimator 1 $2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\hat{\rho}_j$. But this estimator contains so many estimated autocorrelations tha is inconsistent. Intuitively, because each of the estimated autocorrelations conta estimation error, by estimating so many autocorrelations the estimation error this estimator of f_T remains large even in large samples. At the other extreme, o could imagine using only a few sample autocorrelations, for example only the fi sample autocorrelation, and ignoring all the higher autocorrelations. Althou this estimator eliminates the problem of estimating too many autocorrelations has a different problem: it is inconsistent because it ignores the additional aut correlations that appear in Equation (13.13). In short, using too many samp autocorrelations makes the estimator have a large variance, but using too few aut correlations ignores the autocorrelations at higher lags, so in either of the extreme cases the estimator is inconsistent.

Estimators of f_T used in practice strike a balance between these two extrer cases by choosing the number of autocorrelations to include in a way that deper on the sample size T. If the sample size is small, only a few autocorrelations a used, but if the sample size is large, more autocorrelations are included (but s far fewer than T). Specifically, let $\hat{f}_T$ be given by

$$\hat{f}_T = 1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)\widetilde{\rho}_j, \quad (13.1$$

where $\widetilde{\rho}_j = \sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}\Big/\sum_{t=1}^{T}\hat{v}_t^2$, where $\hat{v}_t = (X_t - \overline{X})\hat{u}_t$ (as in the definition of $\hat{\sigma}^2_{\hat{\beta}_1}$). T parameter m in Equation (13.16) is called the **truncation parameter** of the HA estimator because the sum of autocorrelations is shortened, or truncated, include only $m-1$ autocorrelations instead of the $T-1$ autocorrelations appea ing in the population formula in Equation (13.13).

For $\hat{f}_T$ to be consistent, m must be chosen so that it is large in large sampl although still much less than T. One guideline for choosing m in practice is to u the formula

$$m = 0.75T^{1/3}, \quad (13.1$$

rounded to an integer. This formula, which is based on the assumption that there is a moderate amount of autocorrelation in v_t, gives a benchmark rule for determining m as a function of the number of observations in the regression.[1]

The value of the truncation parameter m resulting from Equation (13.17) can be modified using your knowledge of the series at hand. If there is a great deal of serial correlation in v_t, then you could increase m beyond the value from Equation (13.17). On the other hand if v_t has little serial correlation, you could decrease m. Because of the ambiguity associated with the choice of m, it is good practice to try one or two alternative values of m for at least one specification to make sure your results are not sensitive to m.

The HAC estimator in Equation (13.15), with $\hat{f}_T$ given in Equation (13.16), is called the **Newey-West variance estimator**, after the econometricians Whitney Newey and Kenneth West who proposed it. They showed that, when used along with a rule like that in Equation (13.17), under general assumptions this estimator is a consistent estimator of the variance of $\hat{\beta}_1$ (Newey and West, 1987). Their proofs (and those in Andrews (1991)) assume that v_t has more than four moments, which in turn is implied by X_t and u_t having more than eight moments, and this is the reason that the third assumption in Key Concept 13.2 is that X_t and u_t have more than eight moments.

Other HAC estimators. The Newey-West variance estimator is not the only HAC estimator. For example, the weights $(m - j)/m$ in Equation (13.16) can be replaced by different weights. If different weights are used, then the rule for choosing the truncation parameter in Equation (13.17) no longer applies and a different rule, developed for those weights, should be used instead. Discussion of HAC estimators using other weights goes beyond the scope of this book. For more information on this topic, see Hayashi (2000, Section 6.6).

Extension to multiple regression. All the issues discussed in this section generalize to the distributed lag regression model in Key Concept 13.1 with multiple lags and, more generally, to the multiple regression model with serially correlated errors. In particular, if the error term is serially correlated, then the usual OLS standard errors are an unreliable basis for inference and HAC standard errors should be used instead. If the HAC variance estimator used is the Newey-West estimator (the HAC variance estimator based on the weights $(m - j)/m$), then the

[1]Equation (13.17) gives the "best" choice of m if u_t and X_t are first order autoregressive processes with first autocorrelation coefficients 0.5, where "best" means the estimator that minimizes $E(\tilde{\sigma}^2_{\hat{\beta}_1} - \sigma^2_{\hat{\beta}_1})^2$. Equation (13.17) is based on a more general formula derived by Andrews (1991, Equation (5.3)).

HAC Standard Errors

Key Concep 13.3

The problem: The error term u_t in the distributed lag regression model in Key Concept 13.1 can be serially correlated. If so, the OLS coefficient estimators are consistent but in general the usual OLS standard errors are not, resulting in misleading hypothesis tests and confidence intervals.

The solution: Standard errors should be computed using a heteroskedasticity- and autocorrelation-consistent (HAC) estimator of the variance. The HAC estimator involves estimates of $m - 1$ autocovariances as well as the variance; in the case of a single regressor, the relevant formulas are given in Equations (13.15) and (13.16).

In practice, using HAC standard errors entails choosing the truncation parameter m. To do so, use the formula in Equation (13.17) as a benchmark, then increase or decrease m depending on whether your regressors and errors have high or low serial correlation.

truncation parameter m can be chosen according to the rule in Equation (13.17 whether there is a single regressor or multiple regressors. The formula for HAC standard errors in multiple regression is incorporated into modern regression soft ware designed for use with time series data. Because this formula involves matri algebra, we omit it here, and instead refer the reader to Hayashi (2000, Sectio 6.6) for the mathematical details.

HAC standard errors are summarized in Key Concept 13.3.

13.5 Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors

When X_t is strictly exogenous, two alternative estimators of dynamic causal effect are available. The first such estimator involves estimating an autoregressive distrib uted lag (ADL) model instead of a distributed lag model, and calculating the dynami multipliers from the estimated ADL coefficients. This method can entail estimatin fewer coefficients than OLS estimation of the distributed lag model, thus potentiall reducing estimation error. The second method is to estimate the coefficients of th distributed lag model, using **generalized least squares (GLS)** instead of OLS

Although the same number of coefficients in the distributed lag model are estimated by GLS as by OLS, the GLS estimator has a smaller variance. To keep the exposition simple, these two estimation methods are initially laid out and discussed in the context of a distributed lag model with a single lag and AR(1) errors. The potential advantages of these two estimators are greatest, however, when many lags appear in the distributed lag model, so these estimators are then extended to the general distributed lag model with higher order autoregressive errors.

The Distributed Lag Model with AR(1) Errors

Suppose that the causal effect on Y of a change in X lasts for only two periods, that is, it has an initial impact effect β_1 and an effect in the next period of β_2, but no effect thereafter. Then the appropriate distributed lag regression model is the distributed lag model with only current and past values of X_{t-1}:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t. \tag{13.18}$$

As discussed in Section 13.2, in general the error term u_t in Equation (13.18) is serially correlated. One consequence of this serial correlation is that, if the distributed lag coefficients are estimated by OLS, then inference based on the usual OLS standard errors can be misleading. For this reason, Sections 13.3 and 13.4 emphasized the use of HAC standard errors when β_1 and β_2 in Equation (13.18) are estimated by OLS.

In this section, we take a different approach towards the serial correlation in u_t. This approach, which is possible if X_t is strictly exogenous, involves adopting an autoregressive model for the serial correlation in u_t, then using this AR model to derive some estimators that can be more efficient than the OLS estimator in the distributed lag model.

Specifically, suppose that u_t follows the AR(1) model

$$u_t = \phi_1 u_{t-1} + \tilde{u}_t, \tag{13.19}$$

where ϕ_1 is the autoregressive parameter, $\tilde{u}_t$ is serially uncorrelated, and where no intercept is needed because $E(u_t) = 0$. Equations (13.18) and (13.19) imply that the distributed lag model with a serially correlated error can be rewritten as an autoregressive distributed lag model with a serially uncorrelated error. To do so, lag each side of Equation (13.18) and subtract ϕ_1 times this lag from each side:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} &= (\beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t) - \phi_1(\beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_{t-1}) \\ &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} - \phi_1\beta_0 - \phi_1\beta_1 X_{t-1} - \phi_1\beta_2 X_{t-2} + \tilde{u}_t, \end{aligned} \tag{13.20}$$

where the second equality uses $\tilde{u}_t = u_t - \phi_1 u_{t-1}$. Collecting terms in Equ
(13.20), we have that

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \tilde{u}_t \quad (1$$

where

$$\alpha_0 = \beta_0(1 - \phi_1),\ \delta_0 = \beta_1,\ \delta_1 = \beta_2 - \phi_1\beta_1, \text{ and } \delta_2 = -\phi_1\beta_2, \quad (1$$

where β_0, β_1, and β_2 are the coefficients in Equation (13.18) and ϕ_1 is the
correlation coefficient in Equation (13.19).

Equation (13.21) is an ADL model that includes a contemporaneous val
X and two of its lags. We will refer to (13.21) as the ADL representation of th
tributed lag model with autoregressive errors given in Equations (13.18) and (1

The terms in Equation (13.20) can be reorganized differently to obta
expression that is equivalent to Equations (13.21) and (13.22). Let $\tilde{Y}_t = Y_t -$
be the **quasi-difference** of Y_t ("quasi" because it is not the first difference
difference between Y_t and Y_{t-1}; rather, it is the difference between Y_t and ϕ_1
Similarly, let $\tilde{X}_t = X_t - \phi_1 X_{t-1}$ be the quasi-difference of X_t. Then Equation (1
can be written

$$\tilde{Y}_t = \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \tilde{u}_t. \quad (1$$

We will refer to Equation (13.23) as the quasi-difference representati
the distributed lag model with autoregressive errors given in Equations (1
and (13.19).

The ADL model Equation (13.21) (with the parameter restrictions in E
tion (13.22)) and the quasi-difference model in Equation (13.23) are equiva
In both models, the error term, $\tilde{u}_t$, is serially uncorrelated. The two repres
tions, however, suggest different estimation strategies. But before discussing
strategies, we turn to the assumptions under which they yield consistent est
tors of the dynamic multipliers, β_1 and β_2.

The conditional mean zero assumption in the ADL(2,1) and qu differenced models. Because Equations (13.21) (with the restrictions in E
tion (13.22)) and (13.23) are equivalent, the conditions for their estimatio
the same, so for convenience we consider Equation (13.23).

The quasi-difference model in Equation (13.23) is a distributed lag m
involving the quasi-differenced variables with a serially uncorrelated e

Accordingly, the conditions for OLS estimation of the coefficients in Equation (13.23) are the least squares assumptions for the distributed lag model in Key Concept 13.2, expressed in terms of $\tilde{u}_t$ and $\tilde{X}_t$. The critical assumption here is the first assumption which, applied to Equation (13.23), is that $\tilde{X}_t$ is exogenous; that is,

$$E(\tilde{u}_t \mid \tilde{X}_t, \tilde{X}_{t-1}, \ldots) = 0, \tag{13.24}$$

where letting the conditional expectation depend on distant lags of $\tilde{X}_t$ ensures that no additional lags of $\tilde{X}_t$, other than those appearing in Equation (13.23), enter the population regression function.

Because $\tilde{X}_t = X_t - \phi_1 X_{t-1}$, so $X_t = \tilde{X}_t + \phi_1 X_{t-1}$, conditioning on $\tilde{X}_t$ and all of its lags is equivalent to conditioning on X_t and all of its lags. Thus, the conditional expectation condition in Equation (13.24) is equivalent to the condition that $E(\tilde{u}_t \mid X_t, X_{t-1}, \ldots) = 0$. Furthermore, because $\tilde{u}_t = u_t - \phi_1 u_{t-1}$, this condition in turn implies

$$\begin{aligned} 0 &= E(\tilde{u}_t \mid X_t, X_{t-1}, \ldots) \\ &= E(u_t - \phi_1 u_{t-1} \mid X_t, X_{t-1}, \ldots) \\ &= E(u_t \mid X_t, X_{t-1}, \ldots) - \phi_1 E(u_{t-1} \mid X_t, X_{t-1}, \ldots). \end{aligned} \tag{13.25}$$

For the equality in Equation (13.25) to hold for general values of ϕ_1, it must be the case that both $E(u_t \mid X_t, X_{t-1}, \ldots) = 0$ and $E(u_{t-1} \mid X_t, X_{t-1}, \ldots) = 0$. By shifting the time subscripts, the condition that $E(u_{t-1} \mid X_t, X_{t-1}, \ldots) = 0$ can be rewritten as

$$E(u_t \mid X_{t+1}, X_t, X_{t-1}, \ldots) = 0, \tag{13.26}$$

which (by the law of iterated expectations) implies that $E(u_t \mid X_t, X_{t-1}, \ldots) = 0$. In summary, having the zero conditional mean assumption in Equation (13.24) hold for general values of ϕ_1 is equivalent to having the condition in Equation (13.26) hold.

The condition in Equation (13.26) is implied by X_t being strictly exogenous, but it is *not* implied by X_t being (past and present) exogenous. Thus, the least squares assumptions for estimation of the distributed lag model in Equation (13.23) hold if X_t is strictly exogenous, but it is not enough that X_t be (past and present) exogenous.

Because the ADL representation (Equations (13.21) and (13.22)) is equivalent to the quasi-differenced representation (Equation (13.23)), the conditional mean assumption needed to estimate the coefficients of the quasi-differenced representation (that $E(u_t \mid X_{t+1}, X_t, X_{t-1}, \ldots) = 0$) is also the conditional mean assumption for consistent estimation of the coefficients of the ADL representation.

We now turn to the two estimation strategies suggested by these two sentations, estimation of the ADL coefficients and estimation of the coeff of the quasi-differenced model.

OLS Estimation of the ADL Model

The first strategy is to use OLS to estimate the coefficients in the ADL mo Equation (13.21). As the derivation leading to Equation (13.21) shows, i ing the lag of Y and the extra lag of X as regressors makes the error term s uncorrelated (under the assumption that the error follows a first order a gression). Thus the usual OLS standard errors can be used, that is, HAC st errors are not needed when the ADL model coefficients in Equation (13.2 estimated by OLS.

The estimated ADL coefficients are not themselves estimates of the dy multipliers, but the dynamic multipliers can be computed from the ADL cients. A general way to compute the dynamic multipliers is to express th mated regression function as a function of current and past values of X_t, t to eliminate Y_t from the estimated regression function. To do so, repeatedl stitute expressions for lagged values of Y_t into the estimated regression fun Specifically, consider the estimated regression function

$$\hat{Y}_t = \hat{\phi}_1 Y_{t-1} + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2}, \quad ($$

where the estimated intercept has been omitted because it does not ent expression for the dynamic multipliers. Lagging both sides of Equation (13.27) $\hat{Y}_{t-1} = \hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}$, so replacing Y_{t-1} in Equation (13.27) and collecting terms yields

$$\begin{aligned} \hat{Y}_t &= \hat{\phi}_1(\hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}) + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2} \\ &= \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1\hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1) X_{t-2} + \hat{\phi}_1\hat{\delta}_2 X_{t-3} + \hat{\phi}_1^2 Y_{t-2}. \end{aligned} \quad ($$

Repeating this process by repeatedly substituting expressions for Y_{t-2}, Y_{t-} so forth yields

$$\begin{aligned} \hat{Y}_t = {} & \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1\hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-2} + \\ & \hat{\phi}_1(\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-3} + \hat{\phi}_1^2(\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-4} + \cdots. \end{aligned} \quad ($$

The coefficients in Equation (13.29) are the estimators of the dynamic tipliers, computed from the OLS estimators of the coefficients in the ADL in Equation (13.21). If the restrictions on the coefficients in Equation (13.22) were to hold exactly for the *estimated* coefficients, then all the dynamic multipliers beyond the second (that is, the coefficients on X_{t-2}, X_{t-3}, and so forth) would all be zero.[2] However, under this estimation strategy those restrictions will not hold exactly, so the estimated multipliers beyond the second in Equation (13.29) will generally be nonzero.

GLS Estimation

The second strategy for estimating the dynamic multipliers when X_t is strictly exogenous is to use generalized least squares (GLS), which entails estimating Equation (13.23). To describe the GLS estimator, we initially assume that ϕ_1 is known; because in practice it is unknown, this estimator is infeasible, so it is called the infeasible GLS estimator. The infeasible GLS estimator, however, can be modified using an estimator of ϕ_1, which yields a feasible version of the GLS estimator.

Infeasible GLS. Suppose that ϕ_1 were known; then the quasi-differenced variables $\tilde{X}_t$ and $\tilde{Y}_t$ could be computed directly. As discussed in the context of Equations (13.24) and (13.26), if X_t is strictly exogenous, then $E(\tilde{u}_t | \tilde{X}_t, \tilde{X}_{t-1}, \ldots) = 0$. Thus, if X_t is strictly exogenous and if ϕ_1 is known, the coefficients α_0, β_1, and β_2 in Equation (13.23) can be estimated by the OLS regression of $\tilde{Y}_t$ on $\tilde{X}_t$ and $\tilde{X}_{t-1}$ (including an intercept). The resulting estimators of β_1 and β_2—that is, the OLS estimators of the slope coefficients in Equation (13.23) when ϕ_1 is known—are the **infeasible GLS estimators**. This estimator is infeasible because ϕ_1 is unknown, so $\tilde{X}_t$ and $\tilde{Y}_t$ cannot be computed and thus these OLS estimators cannot actually be computed.

Feasible GLS. The **feasible GLS estimator** modifies the infeasible GLS estimator by using a preliminary estimator of ϕ_1, $\hat{\phi}_1$, to compute the estimated quasi-differences. Specifically, the feasible GLS estimators of β_1 and β_2 are the OLS estimators of β_1 and β_2 in Equation (13.23), computed by regressing $\hat{\tilde{Y}}_t$ on $\hat{\tilde{X}}_t$ and $\hat{\tilde{X}}_{t-1}$ (with an intercept), where $\hat{\tilde{X}}_t = X_t - \hat{\phi}_1 X_{t-1}$ and $\hat{\tilde{Y}}_t = Y_t - \hat{\phi}_1 Y_{t-1}$.

The preliminary estimator, $\hat{\phi}_1$, can be computed by first estimating the distributed lag regression in Equation (13.18) by OLS, then using OLS to estimate ϕ_1 in Equation (13.19) with the OLS residuals $\hat{u}_t$ replacing the unobserved regression errors u_t. This version of the GLS estimator is called the Cochrane-Orcutt (1949) estimator.

[2]Substitute the equalities in Equation (13.22) to show that, if those equalities hold, then $\delta_2 + \phi_1\delta_1 + \phi_1^2\delta_0 = 0$.

An extension of the Cochrane-Orcutt method is to continue this proce
atively: use the GLS estimator of β_1 and β_2 to compute revised estimator
use these new residuals to re-estimate ϕ_1; use this revised estimator of ϕ_1 to
pute revised estimated quasi-differences; use these revised estimated quasi-diffe
to re-estimate β_1 and β_2; and continue this process until the estimators of
β_2 converge. This is referred to as the iterated Cochrane-Orcutt estimator

A nonlinear least squares interpretation of the GLS estimator
equivalent interpretation of the GLS estimator is that it estimates the ADL
in Equation (13.21), imposing the parameter restrictions in Equation (1
These restrictions are nonlinear functions of the original parameters β_0,
and ϕ_1, so this estimation cannot be performed using OLS. Instead, the pa
ters can be estimated by nonlinear least squares (NLLS). As discussed in S
9.3, NLLS minimizes the sum of squared mistakes made by the estimated
sion function, recognizing that the regression function is a nonlinear func
the parameters being estimated. In general, NLLS estimation can require s
ticated algorithms for minimizing nonlinear functions of unknown paran
In the special case at hand, however, those sophisticated algorithms are not n
rather, the NLLS estimator can be computed using the algorithm described
for the iterated Cochrane-Orcutt estimator. Thus, the iterated Cochrane-
GLS estimator is in fact the NLLS estimator of the ADL coefficients, sub
the nonlinear constraints in Equation (13.22).

Efficiency of GLS. The virtue of the GLS estimator is that when X is
exogenous and the transformed errors $\tilde{u}_t$ are homoskedastic, it is efficient
linear estimators, at least in large samples. To see this, first consider the inf
GLS estimator. If $\tilde{u}_t$ is homoskedastic, if ϕ_1 is known (so that $\tilde{X}_t$ and $\tilde{Y}_t$ can be
as if they are observed), and if X_t is strictly exogenous, then the Gauss-Marko
orem implies that the OLS estimator of α_0, β_1, and β_2 in Equation (13.23)
cient among all linear conditionally unbiased estimators; that is, the OLS est
of the coefficients in Equation (13.23) is the best linear unbiased estimator, or
(Section 4.9). Because the OLS estimator of Equation (13.23) is the infeasib
estimator, this means that the infeasible GLS estimator is BLUE. The feasib
estimator is similar to the infeasible GLS estimator, except that ϕ_1 is esti
Because the estimator of ϕ_1 is consistent and its variance is inversely proporti
T, the feasible and infeasible GLS estimators have the same variances in larg
ples. In this sense, if X is strictly exogenous, then the feasible GLS estimator is
in large samples. In particular, if X is strictly exogenous then GLS is more e
than the OLS estimator of the distributed lag coefficients discussed in Sectio

The Cochrane-Orcutt and iterated Cochrane-Orcutt estimators presented here are special cases of GLS estimation. In general, GLS estimation involves transforming the regression model so that the errors are homoskedastic and serially uncorrelated, then estimating the coefficients of the transformed regression model by OLS. In general, the GLS estimator is consistent and BLUE if X is strictly exogenous, but is not consistent if X is only (past and present) exogenous. The mathematics of GLS involve matrix algebra, so they are postponed to Section 16.6.

The Distributed Lag Model with Additional Lags and AR(p) Errors

The foregoing discussion of the distributed lag model in Equations (13.18) and (13.19), which has a single lag of X_t and an AR(1) error term, carries over to the general distributed lag model with multiple lags and an AR(p) error term.

The general distributed lag model with autoregressive errors. The general distributed lag model with r lags and AR(p) error term is

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (13.30)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \tilde{u}_t, \quad (13.31)$$

where $\beta_1, \ldots, \beta_{r+1}$ are the dynamic multipliers and $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients of the error term. Under the AR(p) model for the errors, $\tilde{u}_t$ is serially uncorrelated.

Algebra of the sort that led to the ADL model in Equation (13.21) shows that Equations (13.30) and (13.31) imply that Y_t can be written in ADL form:

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \tilde{u}_t, \quad (13.32)$$

where $q = r + p$ and $\delta_0, \ldots, \delta_q$ are functions of the β's and ϕ's in Equations (13.30) and (13.31). Equivalently, the model of Equations (13.30) and (13.31) can be written in quasi-difference form as

$$\tilde{Y}_t = \alpha_0 + \beta_1 \tilde{X}_t + \beta_2 \tilde{X}_{t-1} + \cdots + \beta_{r+1} \tilde{X}_{t-r} + \tilde{u}_t, \quad (13.33)$$

where $\tilde{Y}_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$ and $\tilde{X}_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$.

Conditions for estimation of the ADL coefficients. The foregoing discussion of the conditions for consistent estimation of the ADL coefficients in the

AR(1) case extends to the general model with AR(p) errors. The conditio mean zero assumption for Equation (13.33) is that

$$E(\tilde{u}_t \mid \tilde{X}_t, \tilde{X}_{t-1}, \ldots) = 0. \quad (13.$$

Because $\tilde{u}_t = u_t - \phi_1 u_{t-1} - \phi_2 u_{t-2} - \cdots - \phi_p u_{t-p}$ and $\tilde{X}_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X$ this condition is equivalent to

$$E(u_t \mid X_t, X_{t-1}, \ldots) - \phi_1 E(u_{t-1} \mid X_t, X_{t-1}, \ldots) - \cdots - \phi_p E(u_{t-p} \mid X_t, X_{t-1}, \ldots) = \quad (13.$$

For Equation (13.35) to hold for general values of $\phi_1, \ldots, \phi_p$, it must be case that each of the conditional expectations in Equation (13.35) is zero; equ alently, it must be the case that

$$E(u_t \mid X_{t+p}, X_{t+p-1}, X_{t+p-2}, \ldots) = 0 \quad (13.$$

This condition is not implied by X_t being (past and present) exogenous, it is implied by X_t being strictly exogenous. In fact, in the limit when p is infi (so that the error term in the distributed lag model follows an infinite or autoregression), then the condition in Equation (13.36) becomes the condit in Key Concept 13.1 for strict exogeneity.

Estimation of the ADL model by OLS. As in the distributed lag model v a single lag and an AR(1) error term, the dynamic multipliers can be estima from the OLS estimators of the ADL coefficients in Equation (13.32). The g eral formulas are similar to, but more complicated than, those in Equation (13 and are best expressed using lag multiplier notation; these formulas are give Appendix 13.2. In practice, modern regression software designed for time se regression analysis does these computations for you.

Estimation by GLS. Alternatively, the dynamic multipliers can be e mated by (feasible) GLS. This entails OLS estimation of the coefficients of quasi-differenced specification in Equation (13.33), using estimated quasi- ferences. The estimated quasi-differences can be computed using prelimi estimators of the autoregressive coefficients $\phi_1, \ldots, \phi_p$, as in the AR(1) c The GLS estimator is asymptotically BLUE, in the sense discussed above the AR(1) case.

Key Concept 13.4

Estimation of Dynamic Multipliers Under Strict Exogeneity

The general distributed lag model with r lags and AR(p) error term is

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t \quad (13.37)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \tilde{u}_t. \quad (13.38)$$

If X_t is strictly exogenous, then the dynamic multipliers $\beta_1, \ldots, \beta_{r+1}$ can be estimated by first using OLS to estimate the coefficients of the ADL model

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \tilde{u}_t, \quad (13.39)$$

where $q = r + p$, then computing the dynamic multipliers using regression software. Alternatively, the dynamic multipliers can be estimated by estimating the distributed lag coefficients in Equation (13.37) by GLS.

Estimation of dynamic multipliers under strict exogeneity is summarized in Key Concept 13.4.

Which to use: OLS or GLS? The two estimation options, OLS estimation of the ADL coefficients and GLS estimation of the distributed lag coefficients, have advantages and disadvantages.

The advantage of the ADL approach is that it can reduce the number of parameters needed for estimating the dynamic multipliers, compared to OLS estimation of the distributed lag model. For example, the estimated ADL model in Equation (13.27) led to the infinitely long estimated distributed lag representation in Equation (13.29). To the extent that distributed lag model with only r lags is really an approximation to a longer-lagged distributed lag model, the ADL model thus can provide a simple way to estimate those many longer lags using only a few unknown parameters. Thus, in practice it might be possible to estimate the ADL model in Equation (13.39) with values of p and q much smaller than the value of r needed for OLS estimation of the distributed lag coefficients in Equation (13.37). In other words, the ADL specification can provide a compact, or parsimonious, summary of a long and complex distributed lag (see Appendix 13.2 for additional discussion).

The advantage of the GLS estimator is that, for a given lag length r in t[...] tributed lag model, the GLS estimator of the distributed lag coefficients is [...] efficient than the OLS estimator, at least in large samples. In practice, the [...] advantage of using the ADL approach arises because the ADL specificatio[...] permit estimating fewer parameters than are estimated by GLS.

13.6 Orange Juice Prices and Cold Weather

This section uses the tools of time series regression to squeeze addi[...] insights from our data on Florida temperatures and orange juice prices. [...] how long lasting is the effect of a freeze on the price? Second, has this dy[...] effect been stable or has it changed over the 51 years spanned by the da[...] if so, how?

We begin this analysis by estimating the dynamic causal effects usi[...] method of Section 13.3, that is, by OLS estimation of the coefficients of [...] tributed lag regression of the percentage change in prices ($\%ChgP_t$) on the [...] ber of freezing degree days in that month (FDD_t) and its lagged values. F[...] distributed lag estimator to be consistent, FDD must be (past and present) e[...] nous. As discussed in Section 13.2, this assumption is reasonable here. Hu[...] cannot influence the weather, so treating the weather as if it were ranc[...] assigned experimentally is appropriate. Because FDD is exogenous, we ca[...] mate the dynamic causal effects by OLS estimation of the coefficients in th[...] tributed lag model of Equation (13.4) in Key Concept 13.1.

As discussed in Sections 13.3 and 13.4, the error term can be serially c[...] lated in distributed lag regressions, so it is important to use HAC standard e[...] which adjust for this serial correlation. For the initial results, the truncation [...] meter for the Newey-West standard errors (m in the notation of Section 13.4[...] chosen using the rule in Equation (13.17): because there are 612 monthly c[...] vations, according to that rule $m = 0.75T^{1/3} = 0.75 \times 612^{1/3} = 6.37$, but be[...] m must be an integer this was rounded up to $m = 7$; the sensitivity of the sta[...] errors to this choice of truncation parameter is investigated below.

The results of OLS estimation of the distributed lag regression of $\%Ch$[...] $FDD_t, FDD_{t-1}, \ldots, FDD_{t-18}$ are summarized in column (1) of Table 13.1[...] coefficients of this regression (only some of which are reported in the tabl[...] estimates of the dynamic causal effect on orange juice price changes (in pe[...] for the first 18 months following a unit increase in the number of freezing d[...] days in a month. For example, a single freezing degree day is estimated to in[...]

prices by 0.50% over the month in which the freezing degree day occurs. The subsequent effect on price in later months of a freezing degree day is less: after one month the estimated effect is to increase the price by a further 0.17%, and after two months the estimated effect is to increase the price by an additional .07%. The R^2 from this regression is 0.12, indicating that much of the monthly variation in orange juice prices is not explained by current and past values of *FDD*.

Plots of dynamic multipliers can convey information more effectively than tables such as Table 13.1. The dynamic multipliers from column (1) of Table 13.1 are plotted in Figure 13.2a along with their 95% confidence intervals, computed as the estimated coefficient ±1.96 HAC standard errors. After the initial sharp price rise, subsequent price rises are less, although prices are estimated to rise slightly in each of the first six months after the freeze. As can be seen from Figure 13.2a, for months other than the first the dynamic multipliers are not statistically significantly different from zero at the 5% significance level, although they are estimated to be positive through the seventh month.

Column (2) of Table 13.1 contains the cumulative dynamic multipliers for this specification, that is, the cumulative sum of the dynamic multipliers reported in column (1). These dynamic multipliers are plotted in Figure 13.2b along with their 95% confidence intervals. After one month, the cumulative effect of the freezing degree day is to increase prices by 0.67%, after two months the price is estimated to have risen by 0.74%, and after six months the price is estimated to have risen by 0.90%. As can be seen in Figure 13.2b, these cumulative multipliers increase through the seventh month, because the individual dynamic multipliers are positive for the first seven months. In the eighth month, the dynamic multiplier is negative, so the price of orange juice begins to fall slowly from its peak. After 18 months, the cumulative increase in prices is only 0.37%, that is, the long-run cumulative dynamic multiplier is only 0.37%. This long-run cumulative dynamic multiplier is not statistically significantly different from zero at the 10% significance level ($t = 0.37/0.30 = 1.23$).

Sensitivity analysis. As in any empirical analysis, it is important to check if these results are sensitive to changes in the details of the empirical analysis. We therefore examine three aspects of this analysis: sensitivity to the computation of the HAC standard errors; an alternative specification that investigates potential omitted variable bias; and an analysis of the stability over time of the estimated multipliers.

First, we investigate whether the standard errors reported in the second column of Table 13.1 are sensitive to different choices of the HAC truncation parameter m. In column (3), results are reported for $m = 14$, twice the value used in column (2). The regression specification is the same as in column (2), so the esti-

TABLE 13.1 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice: Selected Estimated Dynamic Multipliers and Cumulative Dynamic Multipliers

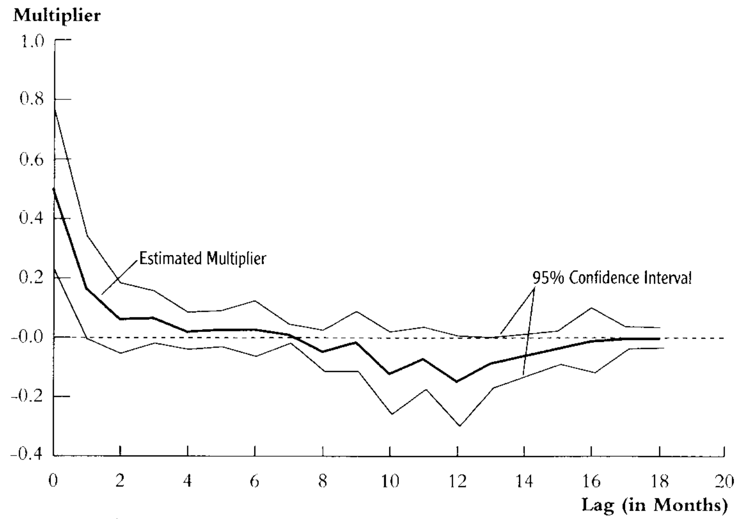
Lag number	(1) Dynamic Multipliers	(2) Cumulative Multipliers	(3) Cumulative Multipliers	(4) Cumulative Multiplie
0	0.50 (0.14)	0.50 (0.14)	0.50 (0.14)	0.51 (0.15)
1	0.17 (0.09)	0.67 (0.14)	0.67 (0.13)	0.70 (0.15)
2	0.07 (0.06)	0.74 (0.17)	0.74 (0.16)	0.76 (0.18)
3	0.07 (0.04)	0.81 (0.18)	0.81 (0.18)	0.84 (0.19)
4	0.02 (0.03)	0.84 (0.19)	0.84 (0.19)	0.87 (0.20)
5	0.03 (0.03)	0.87 (0.19)	0.87 (0.19)	0.89 (0.20)
6 ⋮	0.03 (0.05)	0.90 (0.20)	0.90 (0.21)	0.91 (0.21)
12 ⋮	−0.14 (0.08)	0.54 (0.27)	0.54 (0.28)	0.54 (0.28)
18	0.00 (0.02)	0.37 (0.30)	0.37 (0.31)	0.37 (0.30)
Monthly indicators?	No	No	No	Yes $F = 1.01$ ($p = 0.43$)
HAC standard error truncation parameter (m)	7	7	14	7

All regressions were estimated by OLS using monthly data (described in Appendix 13.1) from January 1950 to December 20 for a total of $T = 612$ monthly observations. The dependent variable is the monthly percentage change in the price of orai juice ($\%ChgP_t$). Regression (1) is the distributed lag regression with the monthly number of freezing degree days and eight of its lagged values, that is, $FDD_t, FDD_{t-1}, \ldots, FDD_{t-18}$, and the reported coefficients are the OLS estimates of the dynai multipliers. The cumulative multipliers are the cumulative sum of estimated dynamic multipliers. All regressions include intercept, which is not reported. Newey-West HAC standard errors, computed using the truncation number given in the f row, are reported in parentheses.

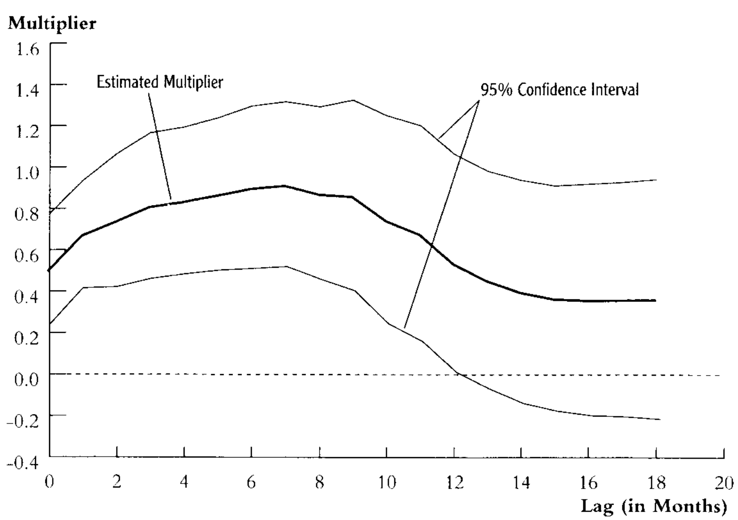
mated coefficients and dynamic multipliers are identical; only the standard erro differ but, as it happens, not by much. We conclude that the results are insen tive to changes in the HAC truncation parameter.

Second, we investigate a possible source of omitted variable bias. Freezes Florida are not randomly assigned throughout the year, but rather occur in t

FIGURE 13.2 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice



(a) Estimated Dynamic Multipliers and 95% Confidence Interval



(b) Estimated Cumulative Dynamic Multipliers and 95% Confidence Interval

The estimated dynamic multipliers show that a freeze leads to an immediate increase in prices. Future price rises are much smaller than the initial impact. The cumulative multiplier shows that freezes have a persistent effect on the level of orange juice prices, with prices peaking seven months after the freeze.

winter (of course). If demand for orange juice is seasonal (is demand fo juice greater in the winter than the summer?), then the seasonal patterns i juice demand could be correlated with *FDD*, resulting in omitted varia The quantity of oranges sold for juice is endogenous: prices and quant simultaneously determined by the forces of supply and demand. Thus, as c in Section 7.2, including quantity would lead to simultaneity bias. Neve the seasonal component of demand can be captured by including seaso ables as regressors. The specification in column (4) of Table 13.1 therefore eleven monthly binary variables, one indicating whether the month is one indicating February, and so forth (as usual one binary variable must b ted to prevent perfect multicollinearity with the intercept). These mont cator variables are not jointly statistically significant at the 10% level ($p =$. the estimated cumulative dynamic multipliers are essentially the same a specifications excluding the monthly indicators. In summary, seasonal fluc in demand are not an important source of omitted variable bias.

Have the dynamic multipliers been stable over time?[3] To assess the of the dynamic multipliers, we need to check whether the distributed lag re coefficients have been stable over time. Because we do not have a specific br in mind, we test for instability in the regression coefficients using the Quan lihood Ratio (QLR) statistic (Key Concept 12.9). The QLR statistic (with 1 ming and HAC variance estimator), computed for the regression of column all coefficients interacted, has a value of 9.08, with $q = 20$ degrees of freed coefficients on FDD_t, its 18 lags, and the intercept). The 1% critical value 12.5 is 2.43, so the QLR statistic rejects at the 1% significance level. The regressions have 40 regressors, a large number; recomputing them for only (so there are 16 regressors and $q = 8$) also results in rejection at the 1 Thus, the hypothesis that the dynamic multipliers are stable is rejected a significance level.

One way to see how the dynamic multipliers have changed over ti compute them for different parts of the sample. Figure 13.3 plots the e cumulative dynamic multipliers for the first third (1950–1966), midd (1967–1983), and final third (1984–2000) of the sample, computed by separate regressions on each subsample. These estimates show an interes noticeable pattern. In the 1950s and early 1960s, a freezing degree day ha and persistent effect on the price. The magnitude of the effect on price of

[3]The discussion of stability in this subsection draws on material from Section 12.7 and can b if that material has not been covered.

FIGURE 13.3 Estimated Cumulative Dynamic Multipliers from Different Sample Periods

The dynamic effect on orange juice prices of freezes changed significantly over the second half of the twentieth century. A freeze had a larger impact on prices during 1950 to 1966 than later, and the effect of a freeze was less persistent during 1984–2000 than earlier.

ing degree day diminished in the 1970s, although it remained highly persistent. In the late 1980s and 1990s, the short-run effect of a freezing degree day was the same as in the 1970s but it became much less persistent, and was essentially eliminated after a year. These estimates suggest that the dynamic causal effect on orange juice prices of a Florida freeze became smaller and less persistent over the second half of the twentieth century.

ADL and GLS estimates. As discussed in Section 13.5, if the error term in the distributed lag regression is serially correlated and *FDD* is strictly exogenous, it is possible to estimate the dynamic multipliers more efficiently than by OLS estimation of the distributed lag coefficients. Before using either the GLS estimator or the estimator based on the ADL model, however, we need to consider whether *FDD* is in fact strictly exogenous. True, humans cannot affect the weather, but does that mean that the weather is *strictly* exogenous? Does the error term u_t in the distributed lag regression have conditional mean zero, given past, present, and *future* values of *FDD*?

The error term in the population counterpart of the distributed lag regression in column (1) of Table 13.1 is the discrepancy between the price and its population prediction based on the past 18 months of weather. This discrepancy might arise for many reasons, one of which is that traders use forecasts of the weather in Orlando. For example, if an especially cold winter is forecasted, then traders would incorporate

NEWS FLASH: Commodity Traders Send Shivers Through Disney World

Although the weather at Disney World in Orlando, Florida, is usually pleasant, now and then a cold spell can settle in. If you are visiting Disney World on a winter evening, should you bring a warm coat? Some people might check the weather forecast on TV, but those in the know can do better: they can check that day's closing price on the New York orange juice futures market!

The financial economist Richard Roll undertook a detailed study of the relationship between orange juice prices and the weather. Roll (1984) examined the effect on prices of cold weather in Orlando, but he also studied the "effect" of changes in the price of an orange juice futures contract (a contract to buy frozen orange juice concentrate at a specified date in the future) on the weather. Roll used daily data from 1975 to 1981 on the prices of OJ futures contracts traded at the New York Cotton Exchange and on daily and overnight temperatures in Orlando. He found that a rise in the price of the futures contract during the trading day in New York predicted cold weather, in particular a freezing spell, in Orlando over the following night. In fact, the market was so effective in predicting cold weather in Florida price rise during the trading day actually pre forecast errors in the official U.S. gover weather forecasts for that night.

Roll's study is also interesting for what he find: although his detailed weather data exp some of the variation in daily OJ futures most of the daily movements in OJ prices rem unexplained. He therefore suggested that t futures market exhibits "excess volatility," t more volatility than can be attributed to move in fundamentals. Understanding why (and if is excess volatility in financial markets is n important area of research in financial econo

Roll's finding also illustrates the difference be forecasting and estimating dynamic causal e Price changes on the OJ futures market are a predictor of cold weather, but that does not mea commodity traders are so powerful that they ca the temperature to fall. Visitors to Disney might shiver after an OJ futures contract price ri they are not shivering *because* of the price rise— of course, they went short in the OJ futures ma

this into the price so the price would be above its predicted value based on t ulation regression; that is, the error term would be positive. If this forecast rate, then in fact future weather would turn out to be cold. Thus future freezin days would be positive ($X_{t+1} > 0$) when the current price is unusually high so that corr(X_{t+1}, u_t) is positive. Stated more simply, although orange juice trad not influence the weather, they can—and do—predict it (see the box). Conse the error term in the price/weather regression is correlated with future v In other words, *FDD* is exogenous, but if this reasoning is true, it is not strictly nous, and the GLS and ADL estimators will not be consistent estimator dynamic multipliers. These estimators therefore are not used in this applica

13.7 Is Exogeneity Plausible? Some Examples

As in regression with cross-sectional data, the interpretation of the coefficients in a distributed lag regression as causal dynamic effects hinges on the assumption that X is exogenous. If X_t or its lagged values are correlated with u_t, then the conditional mean of u_t will depend on X_t or its lags, in which case X is not (past and present) exogenous. Regressors can be correlated with the error term for several reasons, but with economic time series data a particularly important concern is that there could be simultaneous causality, which (as discussed in Section 10.1) results in endogenous regressors. In Section 13.6, we discussed the assumptions of exogeneity and strict exogeneity of freezing degree days in detail. In this section, we examine the assumption of exogeneity in four other economic applications.

U.S. Income and Australian Exports

The United States is an important source of demand for Australian exports. Precisely how sensitive Australian exports are to fluctuations in U.S. aggregate income could be investigated by regressing Australian exports to the United States against a measure of U.S. income. Strictly speaking, because the world economy is integrated, there is simultaneous causality in this relationship: a decline in Australian exports reduces Australian income, which reduces demand for imports from the United States, which reduces U.S. income. As a practical matter, however, this effect is very small because the Australian economy is much smaller than the U.S. economy. Thus, U.S. income plausibly can be treated as exogenous in this regression.

In contrast, in a regression of European Union exports to the United States against U.S. income the argument for treating U.S. income as exogenous is less convincing because demand by residents of the European Union for American exports constitutes a substantial fraction of the total demand for U.S. exports. Thus a decline in U.S. demand for EU exports would decrease EU income, which in turn would decrease demand for U.S. exports and thus decrease U.S.income. Because of these linkages through international trade, EU exports to the United States and U.S. income are simultaneously determined, so in this regression U.S. income arguably is not exogenous. This example illustrates a more general point that whether a variable is exogenous depends on the context: U.S. income is plausibly exogenous in a regression explaining Australian exports, but not in a regression explaining EU exports.

Oil Prices and Inflation

Ever since the oil price increases of the 1970s, macroeconomists have bee ested in estimating the dynamic effect of an increase in the international crude oil on the U.S. rate of inflation. Because oil prices are set in world in large part by foreign oil-producing countries, initially one might think prices are exogenous. But oil prices are not like the weather: members o set oil production levels strategically, taking many factors, including the the world economy, into account. To the extent that oil prices (or quanti set based on an assessment of current and future world economic con including inflation in the United States, oil prices are endogenous.

Monetary Policy and Inflation

The central bankers in charge of monetary policy need to know the e inflation of monetary policy. Because the main tool of monetary polic short-term interest rate (the "short rate"), this means they need to kr dynamic causal effect on inflation of a change in the short rate. Although t rate is determined by the central bank, it is not set by the central bankers dom (as it would be in an ideal randomized experiment) but rather is set e nously: the central bank determines the short rate based on an assessmen current and future state of the economy, especially including the curr future rates of inflation. The rate of inflation in turn depends on the inte (higher interest rates reduce aggregate demand), but the interest rate dep the rate of inflation, its past, and its (expected) future value. Thus the sh is endogenous and the causal dynamic effect of a change in the short rate o inflation cannot be consistently estimated by an OLS regression of the rate tion on current and past interest rates.

The Phillips Curve

The Phillips curve investigated in Chapter 12 is a regression of the chang rate of inflation against lagged changes in inflation and lags of the unempl rate. Because lags of the unemployment rate happened in the past, one mi tially think that there cannot be feedback from current rates of inflation values of the unemployment rate, so that past values of the unemploym can be treated as exogenous. But past values of the unemployment rate w randomly assigned in an experiment; instead, the past unemployment rate was simultaneously determined with past values of inflation. Because inflation and the unemployment rate are simultaneously determined, the other factors that determine inflation contained in u_t are correlated with past values of the unemployment rate, that is, the unemployment rate is not exogenous. It follows that the unemployment rate is not strictly exogenous, so the dynamic multipliers computed using an empirical Phillips curve (for example, the ADL model in Equation (12.17)) are not consistent estimates of the dynamic causal effect on inflation of a change in the unemployment rate.

13.8 Conclusion

Time series data provide the opportunity to estimate the time path of the effect on Y of a change in X, that is, the dynamic causal effect on Y of a change in X. To estimate dynamic causal effects using a distributed lag regression, however, X must be exogenous, as it would be if it were set randomly in an ideal randomized experiment. If X is not just exogenous but is *strictly* exogenous, then the dynamic causal effects can be estimated using an autoregressive distributed lag model or by GLS.

In some applications, such as estimating the dynamic causal effect on the price of orange juice of freezing weather in Florida, a convincing case can be made that the regressor (freezing degree days) is exogenous; thus the dynamic causal effect can be estimated by OLS estimation of the distributed lag coefficients. Even in this application, however, economic theory suggests that the weather is not strictly exogenous, so the ADL or GLS methods are inappropriate. Moreover, in many relations of interest to econometricians, there is simultaneous causality, so the regressor in these specifications are not exogenous, strictly or otherwise. Ascertaining whether or not the regressor is exogenous (or strictly exogenous) ultimately requires combining economic theory, institutional knowledge, and expert judgment.

Summary

1. Dynamic causal effects in time series are defined in the context of a randomized experiment, where the same subject (entity) receives different randomly assigned treatments at different times. The coefficients in a distributed lag regression of Y on X and its lags can be interpreted as the dynamic causal effects when the time path of X is determined randomly and independently of other factors that influence Y.

2. The variable X is (past and present) exogenous if the conditional mea error u_t in the distributed lag regression of Y on current and past values o not depend on current and past values of X. If in addition the conditional m does not depend on future values of X, then X is strictly exogenous.
3. If X is exogenous, then the OLS estimators of the coefficients in a distrib regression of Y on current and past values of X are consistent estimato dynamic causal effects. In general, the error u_t in this regression is seriall lated, so conventional standard errors are misleading and HAC standar must be used instead.
4. If X is strictly exogenous, then the dynamic multipliers can be estimated estimation of an ADL model or by GLS.
5. Exogeneity is a strong assumption that often fails to hold in economic tin data because of simultaneous causality, and the assumption of strict exog even stronger.

Key Terms

dynamic causal effect (489)
distributed lag model (495)
exogeneity (497)
strict exogeneity (497)
dynamic multiplier (501)
impact effect (501)
cumulative dynamic multiplier (501)
long-run cumulative dynamic multiplier (501)
heteroskedasticity- and autocorrelat consistent (HAC) standard erro
truncation parameter (505)
Newey-West variance estimator (50
generalized least squares (GLS) (507
quasi-difference (509)
infeasible GLS estimator (512)
feasible GLS estimator (512)

Review the Concepts

13.1 In the 1970s a common practice was to estimate a distributed lag mod ing changes in nominal gross domestic product (Y) to current a changes in the money supply (X). Under what assumptions will this sion estimate the causal effects of money on nominal GDP? Ar assumptions likely to be satisfied in a modern economy like the United

13.2 Suppose that X is strictly exogenous. A researcher estimates an AI model, calculates the regression residual, and finds the residual to be serially correlated. Should the researcher estimate a new ADL model with additional lags, or simply use HAC standard errors for the ADL(1,1) estimated coefficients?

13.3 Suppose that a distributed lag regression is estimated, where the dependent variable is ΔY_t instead of Y_t. Explain how you would compute the dynamic multipliers of X_t on Y_t.

13.4 Suppose that you added FDD_{t+1} as an additional regressor in Equation (13.2). If FDD is strictly exogenous, would you expect the coefficient on FDD_{t+1} to be zero or nonzero? Would your answer change if FDD is exogenous but not strictly exogenous?

Exercises

***13.1** Increases in oil prices have been blamed for several recessions in developed countries. To quantify the effect of oil prices on real economic activity researchers have done regressions like those discussed in this chapter. Let GDP_t denote the value of quarterly gross domestic product in the United States, and let $Y_t = 100\ln(GDP_t / GDP_{t-1})$ be the quarterly percentage change in GDP. James Hamilton, an econometrician and macroeconomist, has suggested that oil prices adversely affect that economy only when they jump above their values in the recent past. Specifically, let O_t equal the greater of zero or the percentage point difference between oil prices at date t and their maximum value during the past year. A distributed lag regression relating Y_t and O_t, estimated over 1955:I–2000:IV, is

$$\underset{(0.1)}{\hat{Y}_t = 1.0} - \underset{(0.054)}{0.055O_t} - \underset{(0.057)}{0.026O_{t-1}} - \underset{(0.048)}{0.031O_{t-2}} - \underset{(0.042)}{0.109O_{t-3}} - \underset{(0.053)}{0.128O_{t-4}}$$

$$+ \underset{(0.025)}{0.008O_{t-5}} + \underset{(0.048)}{0.025O_{t-6}} - \underset{(0.039)}{0.019O_{t-7}} + \underset{(0.042)}{0.067O_{t-8}}.$$

a. Suppose that oil prices jump 25% above their previous peak value and stay at this new higher level (so that $O_t = 25$ and $O_{t+1} = O_{t+2} = \cdots = 0$). What is the predicted effect on output growth for each quarter over the next two years?

b. Construct a 95% confidence interval for your answers in (a).

c. What is the predicted cumulative change in GDP growth over eight quarters?

d. The HAC F-statistic testing whether the coefficients on O_t a[...] are zero is 3.49. Are the coefficients significantly different fr[...]

13.2 Macroeconomists have also noticed that interest rates change foll[...] price jumps. Let R_t denote the interest rate on 3-month Treasur[...] percentage points at an annual rate). The distributed lag regressio[...] the change in R_t (ΔR_t) to O_t estimated over 1955:I–2000:IV is

$$\widehat{\Delta R_t} = \underset{(0.06)}{0.07} + \underset{(0.045)}{0.062O_t} + \underset{(0.034)}{0.048O_{t-1}} - \underset{(0.028)}{0.014O_{t-2}} - \underset{(0.169)}{0.086O_{t-3}} - \ldots$$

$$+ \underset{(0.065)}{0.023O_{t-5}} - \underset{(0.047)}{0.010O_{t-6}} - \underset{(0.038)}{0.100O_{t-7}} - \underset{(0.025)}{0.014O_{t-8}}.$$

a. Suppose that oil prices jump 25% above their previous peak [...] stay at this new higher level (so that $O_t = 25$ and $O_{t+1} = O_{t+2} =$ [...] What is the predicted change in interest rates for each quarte[...] the next two years?

b. Construct 95% confidence intervals for your answers to (a).

c. What is the effect of this change in oil prices on the level of i[...] rates in period $t + 8$? How is your answer related to the cum[...] multiplier?

d. The HAC F-statistic testing whether the coefficients on O_t an[...] are zero is 4.25. Are the coefficients significantly different fro[...]

13.3 Consider two different randomized experiments. In experime[...] prices are set randomly and the Central Bank reacts according t[...] policy rules in response to economic conditions, including chan[...] oil price. In experiment B, oil prices are set randomly and th[...] Bank holds interest rates constant, and in particular does not re[...] the oil price changes. In both, GDP growth is observed. Now [...] that oil prices are exogenous in the regression in Exercise 13.1. [...] experiment, A or B, does the dynamic causal effect estimated in [...] 13.1 correspond?

13.4 Suppose that oil prices are strictly exogenous. Discuss how y[...] improve upon the estimates of the dynamic multipliers in Exercis[...]

13.5 Derive Equation (13.7) from Equation (13.4) and show that $\delta_0 =$ [...] β_1, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$ (etc.). (*Hint:* Note that $X_t = \Delta X_t$ [...] $\cdots + \Delta X_{t-p+1} + X_{t-p}$.)

APPENDIX 13.1 The Orange Juice Data Set

The orange juice price data are the frozen orange juice component of processed foods and feeds group of the producer price index (PPI), collected by the U.S. Bureau of Labor Statistics (BLS series wpu02420301). The orange juice price series was divided by the overall PPI for finished goods to adjust for general price inflation. The freezing degree days series was constructed from daily minimum temperatures recorded at Orlando area airports, obtained from the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce. The *FDD* series was constructed so that its timing, and the timing of the orange juice price data, were approximately aligned. Specifically, the frozen orange juice price data are collected by surveying a sample of producers in the middle of every month, although the exact date varies from month to month. Accordingly, the *FDD* series was constructed to be the number of freezing degree days from the 11th of one month to the 10th of the next month; that is, *FDD* is the maximum of zero and 32 minus the minimum daily temperature, summed over all days from the 11th to the 10th. Thus, $\%ChgP_t$ for February is the percentage change in real orange juice prices from mid-January to mid-February, and FDD_t for February is the number of freezing degree days from January 11 to February 10.

APPENDIX 13.2 The ADL Model and Generalized Least Squares in Lag Operator Notation

This appendix presents the distributed lag model in lag operator notation, derives the ADL and quasi-differenced representations of the distributed lag model, and discusses the conditions under which the ADL model can have fewer parameters than the original distributed lag model.

The Distributed Lag, ADL, and Quasi-Differenced Models, in Lag Operator Notation

As defined in Appendix 12.3, the lag operator, L, has the property that $L^jX_t = X_{t-j}$, and the distributed lag $\beta_1X_t + \beta_2X_{t-1} + \cdots + \beta_{r+1}X_{t-r}$ can be expressed as $\beta(L)X_t$, where $\beta(L) = \sum_{j=0}^{r}\beta_{j+1}L^j$, where $L^0 = 1$. Thus, the distributed lag model in Key Concept 13.1 (Equation (13.4)) can be written in lag operator notation as

$$Y_t = \beta_0 + \beta(L)X_t + u_t. \qquad (13.40)$$

In addition, if the error term u_t follows an AR(p), then it can be written a

$$\phi(L)u_t = \tilde{u}_t,$$

where $\phi(L) = \sum_{j=0}^{p}\phi_jL^j$, where $\phi_0 = 1$ and $\tilde{u}_t$ is serially uncorrelated (note that $\phi_1, \ldots$ defined here are the negative of $\phi_1, \ldots, \phi_p$ in the notation of Equation (13.31)

To derive the ADL model, premultiply each side of Equation (13.40) by ϕ(L

$$\phi(L)Y_t = \phi(L)[\beta_0 + \beta(L)X_t + u_t] = \alpha_0 + \delta(L)X_t + \tilde{u}_t,$$

where

$$\alpha_0 = \phi(1)\beta_0 \text{ and } \delta(L) = \phi(L)\beta(L), \text{ where } \phi(1) = \sum_{j=0}^{p}\phi_j.$$

To derive the quasi-differenced model, note that $\phi(L)\beta(L)X_t = \beta(L)\phi(L)X_t$ where $\tilde{X}_t = \phi(L)X_t$. Thus, rearranging Equation (13.42) yields

$$\tilde{Y}_t = \alpha_0 + \beta(L)\tilde{X}_t + \tilde{u}_t,$$

where $\tilde{Y}_t$ is the quasi-difference of Y_t, that is, $\tilde{Y}_t = \phi(L)Y_t$.

The ADL and GLS Estimators

The OLS estimator of the ADL coefficients is obtained by OLS estimation of (13.42). The original distributed lag coefficients are β(L) which, in terms of the coefficients, is $\beta(L) = \delta(L)/\phi(L)$; that is, the coefficients in β(L) satisfy the re implied by $\phi(L)\beta(L) = \delta(L)$. Thus, the estimator of the dynamic multipliers base OLS estimators of the coefficients of the ADL model, $\hat{\delta}(L)$ and $\hat{\phi}(L)$, is

$$\hat{\beta}^{ADL}(L) = \hat{\delta}(L)/\hat{\phi}(L).$$

The expressions for the coefficients in Equation (13.29) in the text are obtained cial case of Equation (13.45) when $r = 1$ and $p = 1$.

The feasible GLS estimator is computed by obtaining a preliminary estimato computing estimated quasi-differences, estimating β(L) in Equation (13.44) us estimated quasi-differences, and (if desired) iterating until convergen iterated GLS estimator is the NLLS estimator computed by NLLS estimatio ADL model in Equation (13.42), subject to the nonlinear restrictions on the pa contained in Equation (13.43).

As stressed in the discussion surrounding Equation (13.36) in the text, it is no for X_t to be (past and present) exogenous to use either of these estimation met exogeneity alone does not ensure that Equation (13.36) holds. If, however, X is strictly exogenous, then Equation (13.36) does hold and, assuming that Assumptions 2–4 of Key Concept 12.6 hold, these estimators are consistent and asymptotically normal. Moreover, the usual (cross-sectional heteroskedasticity-robust) OLS standard errors provide a valid basis for statistical inference.

Parameter reduction using the ADL model. Suppose that the distributed lag polynomial $\beta(L)$ can be written as a ratio of lag polynomials, $\theta_1(L)/\theta_2(L)$, where $\theta_1(L)$ and $\theta_2(L)$ are both lag polynomials of a low degree. Then $\phi(L)\beta(L)$ in Equation (13.43) is $\phi(L)\beta(L) = \phi(L)\theta_1(L)/\theta_2(L) = [\phi(L)/\theta_2(L)]\theta_1(L)$. If it so happens that $\phi(L) = \theta_2(L)$, then $\delta(L) = \phi(L)\beta(L) = \theta_1(L)$. If the degree of $\theta_1(L)$ is low, then q, the number of lags of X_t in the ADL model, can be much less than r. Thus, under these assumptions, estimation of the ADL model entails estimating potentially many fewer parameters than the original distributed lag model. It is in this sense that the ADL model can achieve a more parsimonious parameterizations (that is, use fewer unknown parameters) than the distributed lag model.

As developed here, the assumption that $\phi(L)$ and $\theta_2(L)$ happen to be the same seems like a coincidence that would not occur in an application. However, the ADL model is able to capture a large number of shapes of dynamic multipliers with only a few coefficients. For this reason, unrestricted estimation of the ADL model presents an attractive way to approximate a long distributed lag (that is, many dynamic multipliers) whenever X is strictly exogenous.

CHAPTER 14

Additional Topics in Time Series Regression

This chapter takes up some further topics in time series regression, st with forecasting. Chapter 12 considered forecasting a single variabl practice, however, you might want to forecast two or more variables such rate of inflation and the growth rate of the GDP. Section 14.1 introduces model for forecasting multiple variables, vector autoregressions (VARs), i which lagged values of two or more variables are used to forecast future v of those variables. Chapter 12 also focused on making forecasts one peric (e.g., one quarter) into the future, but making forecasts two, three, or mc periods into the future also is important. Methods for making such foreca discussed in Section 14.2.

Sections 14.3 and 14.4 return to the topic of Section 12.6, stochastic trends. Section 14.3 introduces additional models of stochastic trends and alternative test for a unit autoregressive root. Section 14.4 introduces the concept of cointegration, which arises when two variables share a comm stochastic trend, that is, when each variable contains a stochastic trend, bu weighted difference of the two variables does not.

In some time series data, especially financial data, the variance chang time: sometimes the series exhibits high volatility, while at other times th volatility is low, so that the data exhibit clusters of volatility. Section 14.5 discusses volatility clustering and introduces models in which the variance the forecast error changes over time, that is, models in which the forecast is conditionally heteroskedastic. Models of conditional heteroskedasticity several applications. One application is computing forecast intervals, wher width of the interval changes over time to reflect periods of high or low uncertainty. Another application is to forecasting the uncertainty of returns on an asset, such as a stock, which in turn can be useful in assessing the risk of owning a stock.

14.1 Vector Autoregressions

Chapter 12 focused on forecasting the rate of inflation, but in reality economic forecasters are in the business of forecasting other key macroeconomic variables as well, such as the rate of unemployment, the growth rate of GDP, and interest rates. One approach is to develop a separate forecasting model for each variable using the methods of Section 12.4. Another approach, however, is to develop a single model that can forecast all the variables, which can help to make the forecasts mutually consistent. One way to forecast several variables with a single model is to use a vector autoregression (VAR). A VAR extends the univariate autoregression to multiple time series variables, that is, it extends the univariate autoregression to a "vector" of time series variables.

The VAR Model

A **vector autoregression**, or **VAR**, with two time series variables, Y_t and X_t, consists of two equations: in one, the dependent variable is Y_t; in the other, the dependent variable is X_t. The regressors in both equations are lagged values of both variables. More generally, a VAR with k time series variables consists of k equations, one for each of the variables, where the regressors in all equations are lagged values of all the variables. The coefficients of the VAR are estimated by estimating each of the equations by OLS.

VARs are summarized in Key Concept 14.1.

Inference in VARs. Under the VAR assumptions, the OLS estimators are consistent and have a joint normal distribution in large samples. Accordingly, statistical inference proceeds in the usual manner; for example, 95% confidence intervals on coefficients can be constructed as the estimated coefficient ±1.96 standard errors.

One new aspect of hypothesis testing arises in VARs because a VAR with k variables is a collection, or system, of k equations. Thus it is possible to test joint hypotheses that involve restrictions across multiple equations.

Key Con 14.1

Vector Autoregressions

A vector autoregression (VAR) is a set of k time series regressions, in which the regressors are lagged values of all k series. A VAR extends the univariate autoregression to a list, or "vector," of time series variables. When the number of lags in each of the equations is the same and is equal to p, the system of equations is called a VAR(p).

In the case of two time series variables, Y_t and X_t, the VAR(p) consists of the two equations

$$Y_t = \beta_{10} + \beta_{11}Y_{t-1} + \cdots + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1} + \cdots + \gamma_{1p}X_{t-p} + u_{1t} \qquad (14.1)$$

$$X_t = \beta_{20} + \beta_{21}Y_{t-1} + \cdots + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1} + \cdots + \gamma_{2p}X_{t-p} + u_{2t} \qquad (14.2)$$

where the β's and the γ's are unknown coefficients and u_{1t} and u_{2t} are error terms.

The VAR assumptions are the time series regression assumptions of Key Concept 12.6, applied to each equation. The coefficients of a VAR are estimated by estimating each equation by OLS.

For example, in the two-variable VAR(p) in Equations (14.1) and (14
could ask whether the correct lag length is p or $p - 1$; that is, you co
whether the coefficients on Y_{t-p} and X_{t-p} are zero in these two equations. T
hypothesis that these coefficients are zero is

$$H_0\colon \beta_{1p} = 0, \beta_{2p} = 0, \gamma_{1p} = 0, \text{ and } \gamma_{2p} = 0.$$

The alternative hypothesis is that at least one of these four coefficients is n
Thus the null hypothesis involves coefficients from *both* of the equatio
from each equation.

Because the estimated coefficients have a jointly normal distribution
samples, it is possible to test restrictions on these coefficients by compu
F-statistic. The precise formula for this statistic is complicated because th
tion must handle multiple equations, so we omit it. In practice, most mode
ware packages have automated procedures for testing hypotheses on coef
in systems of multiple equations.

How many variables should be included in a VAR? The number of coefficients in each equation of a VAR is proportional to the number of variables in the VAR. For example, a VAR with five variables and four lags will have 21 coefficients (four lags each of five variables, plus the intercept) in each of the five equations, for a total of 105 coefficients! Estimating all these coefficients increases the amount of estimation error entering a forecast, which can result in a deterioration of the accuracy of the forecast.

The practical implication is that one needs to keep the number of variables in a VAR small and, especially, to make sure that the variables are plausibly related to each other so that they will be useful for forecasting each other. For example, we know from a combination of empirical evidence (such as that discussed in Chapter 12) and economic theory that the inflation rate, the unemployment rate, and the short-term interest rate are related to each other, suggesting that these variables could help to forecast each other in a VAR. Including an unrelated variable in a VAR, however, introduces estimation error without adding predictive content, thereby reducing forecast accuracy.

Determining lag lengths in VARs.[1] Lag lengths in a VAR can be determined using either F-tests or information criteria.

The information criterion for a system of equations extends the single-equation information criterion in Section 12.5. To define this information criterion we need to adopt matrix notation. Let Σ_u be the $k \times k$ covariance matrix of the VAR errors, and let $\hat{\Sigma}_u$ be the estimate of the covariance matrix where the i,j element of $\hat{\Sigma}_u$ is $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}$, where $\hat{u}_{it}$ is the OLS residual from the i^{th} equation and $\hat{u}_{jt}$ is the OLS residual from the j^{th} equation. The BIC for the VAR is

$$\text{BIC}(p) = \ln[\det(\hat{\Sigma}_u)] + k(kp+1)\frac{\ln T}{T}, \qquad (14.4)$$

where $\det(\hat{\Sigma}_u)$ is the determinant of the matrix $\hat{\Sigma}_u$. The AIC is computed using Equation (14.4), modified by replacing the term "$\ln T$" by "2".

The expression for the BIC for the k equations in the VAR in Equation (14.4) extends the expression for a single equation given in Section 12.5. When there is a single equation, the first term simplifies to $\ln(SSR(p)/T)$. The second term in Equation (14.4) is the penalty for adding additional regressors; $k(kp + 1)$ is the total number of regression coefficients in the VAR (there are k equations, each of which has an intercept and p lags of each of the k time series variables).

[1] This section uses matrices and may be skipped for less mathematical treatments.

Lag length estimation in a VAR using the BIC proceeds analogously single equation case: among a set of candidate values of p, the estimated lag $\hat{p}$ is the value of p that minimizes BIC(p).

Using VARs for causal analysis. The discussion so far has focused o VARs for forecasting. Another use of VAR models is for analyzing causal re ships among economic time series variables; indeed, it was for this purpo VARs were first introduced to economics by the econometrician and macr omist Christopher Sims (1980). The use of VARs for causal inference is kn structural VAR modeling, "structural" because in this application VARs are model the underlying structure of the economy. Structural VAR analysis u techniques introduced in this section in the context of forecasting, plus som tional tools. The biggest conceptual difference between using VARs for fore and using them for structural modeling, however, is that structural modeling r very specific assumptions, derived from economic theory and institutional k edge, of what is exogenous and what is not. The discussion of structural V best undertaken in the context of estimation of systems of simultaneous equ which goes beyond the scope of this book. For an introduction to using VA forecasting and policy analysis, see Stock and Watson (2001). For additional ematical detail on structural VAR modeling, see Hamilton (1994) or Watson

A VAR Model of the Rates of Inflation and Unemployment

As an illustration, consider a two-variable VAR for the inflation rate, Inf_t, a rate of unemployment, $Unemp_t$. As in Chapter 12, we treat the rate of infla having a stochastic trend, so that it is appropriate to transform it by compu first difference, ΔInf_t.

A VAR for ΔInf_t and $Unemp_t$ consists of two equations, one in which the dependent variable and one in which $Unemp_t$ is the dependent variabl regressors in both equations are lagged values of ΔInf_t and $Unemp_t$. In Sectic (Equation (12.17)), we reported the following regression of ΔInf_t on four la of ΔInf_t and $Unemp_t$, estimated using quarterly U.S. data from 1962:I–199

$$\widehat{\Delta Inf_t} = \underset{(0.47)}{1.32} - \underset{(0.09)}{0.36}\Delta Inf_{t-1} - \underset{(0.10)}{0.34}\Delta Inf_{t-2} + \underset{(0.08)}{0.07}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4}$$

$$- \underset{(0.47)}{2.68}Unemp_{t-1} + \underset{(0.89)}{3.43}Unemp_{t-2} - \underset{(0.89)}{1.04}Unemp_{t-3} + \underset{(0.44)}{0.07}Unemp_{t-4}.$$

The adjusted R^2 is $\bar{R}^2 = 0.35$.

This is in fact the first equation of a VAR(4) model of the change in inflation and the unemployment rate. The second equation has the same regressors, but the dependent variable is the unemployment rate:

$$\begin{aligned}\widehat{Unemp_t} = &\underset{(0.09)}{0.12} + \underset{(0.020)}{0.043}\Delta Inf_{t-1} + \underset{(0.015)}{0.000}\Delta Inf_{t-2} + \underset{(0.16)}{0.021}\Delta Inf_{t-3} + \underset{(0.15)}{0.021}\Delta Inf_{t-4} \\ &+ \underset{(0.12)}{1.68}Unemp_{t-1} - \underset{(0.20)}{0.70}Unemp_{t-2} - \underset{(0.20)}{0.03}Unemp_{t-3} + \underset{(0.09)}{0.02}Unemp_{t-4}.\end{aligned} \quad (14.6)$$

The adjusted R^2 is $\bar{R}^2 = 0.975$.

Equations (14.5) and (14.6), taken together, are a VAR(4) model of the change in the rate of inflation, ΔInf_t, and the unemployment rate, $Unemp_t$.

These VAR equations can be used to perform Granger causality tests. The F-statistic testing the null hypothesis that the coefficients on $Unemp_{t-1}$, $Unemp_{t-2}$, $Unemp_{t-3}$, and $Unemp_{t-4}$ are zero in the inflation equation (Equation (14.5)) is 8.51, which has a p-value less than 0.001. Thus, the null hypothesis is rejected, so we can conclude that the unemployment rate is a useful predictor of changes in inflation, given lags in inflation (that is, the unemployment rate Granger-causes changes in inflation). Similarly, the F-statistic testing the hypothesis that the coefficients on the four lags of ΔInf_t are zero in the unemployment equation (Equation (14.6)) is 2.41, which has a p-value of 0.051. Thus four lags of the change in the inflation rate Granger-cause the unemployment rate at the 10% significance level but not at the 5% significance level.

Forecasts of the rates of inflation and unemployment one period ahead are obtained exactly as discussed in Section 12.4. The forecast of the change of inflation from 1999:IV to 2000:I, based on Equation (14.5) and using data through 1999:IV, was computed in Section 12.4; this forecast is $\widehat{\Delta Inf}_{2000:I|1999:IV} = 0.5$ percentage points. A similar calculation using Equation (14.6) gives a forecast of the unemployment rate in 2000:I based on data through 1999:IV of $\widehat{Unemp}_{2000:I|1999:IV} = 4.1\%$, very close to its actual value, $Unemp_{2000:I} = 4.0\%$.

14.2 Multiperiod Forecasts

The discussion of forecasting so far has focused on making forecasts one period in advance. Often, however, forecasters are called upon to make forecasts further into the future. The forecasting regression models of Chapter 12 can produce such multiperiod forecasts, but some modifications are needed. This section

discusses those modifications, first for univariate autoregressions and th multivariate forecasting.

Multiperiod Forecasting: Univariate Autoregressions

We present two methods for making multiperiod forecasts from a univ autoregression. The first is the "multiperiod regression method"; the second "iterated autoregression" method.

The multiperiod regression method: AR(1). Suppose you want to autoregression to make a forecast two periods ahead. In the multiperiod r sion method, each predictor is replaced by its lagged value, and the coefficie this modified autoregression are estimated by OLS. If Y_t follows an AR(1) in the one-step ahead regression Y_t is regressed onto a constant and Y_{t-1}. two-step ahead regression, however, Y_{t-1} is unavailable, so the two-step regression entails regressing Y_t onto a constant and Y_{t-2}.

For example, consider forecasting the quarterly change in the inflatio two quarters ahead using an AR(1) model for the change in inflation. The ified two-period ahead regression, estimated over the period 1962:I–1999:

$$\widehat{\Delta Inf}_{t|t-2} = \underset{(0.12)}{0.02} - \underset{(0.09)}{0.30}\Delta Inf_{t-2},$$

where $\widehat{\Delta Inf}_{t|t-2}$ is the predicted value of ΔInf_t based on values of the inflatio through period $t-2$.

Equation (14.7) illustrates the key idea of the multiperiod regression meth data from period $t-1$ appear as a regressor, so only values of inflation dated and earlier are used to forecast ΔInf_t. For example, according to Equation (the forecast of the change of inflation between the first and the second quarter of based on information through the fourth quarter of 1999, is $\widehat{\Delta Inf}_{2000:II|1999}$ $0.02 - 0.30\Delta Inf_{1999:IV}$. From Table 12.1 (p. 434), $\Delta Inf_{1999:IV} = 0.4$. $\widehat{\Delta Inf}_{2000:II|1999:IV} = 0.02 - 0.30 \times 0.4 = -0.1$. That is, based on data throug fourth quarter of 1999, inflation is forecasted to decline by one tenth of a centage point from the first to the second quarter of 2000.

To compute forecasts into the more distant future, the multiperiod re sion method involves using more distant lags. For example, when Y_t follo AR(1), the three-period ahead forecast is computed from a regression of Y_t a constant and Y_{t-3}.

The multiperiod regression method: AR(p). The multiperiod regression approach can be extended to higher order autoregressions by including additional lagged values in the regression. In general, in an AR(p), the modified two-step ahead regression would entail regressing Y_t onto a constant and $Y_{t-2}, Y_{t-3}, \ldots, Y_{t-p-1}$. Similarly, the three-step ahead regression would entail regressing Y_t onto a constant and $Y_{t-3}, Y_{t-4}, \ldots, Y_{t-p-2}$.

For example, the two-period ahead forecast from an AR(4) model for ΔInf_t is obtained using the regression of ΔInf_t onto $\Delta Inf_{t-2}, \ldots, \Delta Inf_{t-5}$:

$$\widehat{\Delta Inf}_{t|t-2} = \underset{(0.10)}{0.02} - \underset{(0.08)}{0.27}\Delta Inf_{t-2} + \underset{(0.09)}{0.25}\Delta Inf_{t-3} - \underset{(0.10)}{0.08}\Delta Inf_{t-4} - \underset{(0.08)}{0.01}\Delta Inf_{t-5}. \quad (14.8)$$

The values in Table 12.1 and the coefficients in Equation (14.8) can be used to forecast the change in inflation from 2000:I to 2000:II: $\widehat{\Delta Inf}_{2000:II|1999:IV} = 0.02 - 0.27\Delta Inf_{1999:IV} + 0.25\Delta Inf_{1999:III} - 0.08\Delta Inf_{1999:II} - 0.01\Delta Inf_{1999:I} = 0.02 - 0.27 \times 0.4 + 0.25 \times 0.0 - 0.08 \times 1.1 - 0.01 \times (-0.4) = -0.2$. That is, based on Equation (14.8), based on inflation data through the fourth quarter of 1999, inflation is forecasted to decline by 0.2 percentage points from the first to the second quarter of 2000.

To make forecasts three periods in advance using an AR(4), Equation (14.8) would be modified so that ΔInf_t is regressed onto $\Delta Inf_{t-3}, \ldots, \Delta Inf_{t-6}$. More generally, to make an h-period ahead forecast of Y_t using an AR(p), the variable of interest is regressed on its p lags, where the most recent date of the regressors is $t - h$.

Standard errors in multiperiod regressions. Because the dependent variable in a multiperiod regression occurs two or more periods into the future, the error term in a multiperiod regression is serially correlated. To see this, consider the two-period ahead inflation forecasts, and suppose there is a surprise jump in oil prices next quarter. Then today's two-period ahead forecast of inflation will be too low because it does not incorporate this unexpected event. Because the oil price rise was also unknown last quarter, the two-period ahead forecast made last quarter will also be too low: thus, the surprise oil price hike next quarter means that *both* last quarter's and this quarter's two-period ahead forecasts are too low. Because of such intervening events, the error term in a multiperiod regression is serially correlated.

As discussed in Section 13.4, if the error term is serially correlated, the usual OLS standard errors are incorrect or, more precisely, they are not a reliable basis for inference. Therefore heteroskedasticity- and autocorrelation-consistent (HAC) standard errors must be used with multiperiod regressions. The standard errors reported in this section for multiperiod regressions therefore are Newey-West HAC

standard errors, where the truncation parameter m is set according to
(13.17); for these data (for which $T = 152$), Equation (13.17) yields m
longer forecast horizons the amount of overlap, and thus the degree of s
relation in the error, increases: in general, the first $h - 1$ autocorrelation co
of the errors in an h-period ahead regression are nonzero. Thus, larger va
than indicated by Equation (13.17) are appropriate for multiperiod regress
long forecast horizons.

The iterated AR forecast method: AR(1). The iterated AR forecas
uses the AR model to extend a one-period ahead forecast to two or m
ods ahead. The two-period ahead forecast is computed in two steps. In
step, the one-period ahead forecast is computed as in Section 12.3. In th
step, the two-period ahead forecast is computed using the one-period ah
cast for the intervening period. Thus, the one-period ahead forecast is u
intermediate step to make the two-period ahead forecast. For more dist
zons, this process is repeated or "iterated."

As an example, consider the first order autoregression for ΔInf_t (
(12.7)), which is

$$\widehat{\Delta Inf}_t = \underset{(0.14)}{0.02} - \underset{(0.11)}{0.21}\Delta Inf_{t-1}.$$

The first step in computing the two-quarter ahead forecast of $\Delta Inf_{2000:II}$
Equation (14.9) using data through 1999:IV is to compute the one-quar
forecast of $\Delta Inf_{2000:I}$ based on data through 1999:IV: $\widehat{\Delta Inf}_{2000:I|1999:IV}$ =
$0.21\Delta Inf_{1999:IV} = 0.02 - 0.21 \times 0.4 = -0.1$. In the second step, this forecast
tuted into Equation (14.9); that is, $\widehat{\Delta Inf}_{2000:II|1999:IV} = 0.02 - 0.21\,\widehat{\Delta Inf}_{2000:I}$
$0.02 - 0.21 \times (-0.1) = 0.0$. Thus, based on information through the fou
ter of 1999, this forecast is that the rate of inflation will not change bet
first and second quarters of 2000.

The iterated AR forecast method: AR(p). The iterated AR(1) s
extended to an AR(p) by replacing Y_{t-1} in the estimated AR(p) with it
made the previous period.

For example, consider the iterated two-step ahead forecast of inflati
on the AR(4) model from Section 12.3 (Equation (12.13)),

$$\widehat{\Delta Inf}_t = \underset{(0.12)}{0.02} - \underset{(0.10)}{0.21}\Delta Inf_{t-1} - \underset{(0.09)}{0.32}\Delta Inf_{t-2} + \underset{(0.09)}{0.19}\Delta Inf_{t-3} - \underset{(0.10)}{0.04}\Delta Inf_{t-4}.$$

Key Concept 14.2

Multiperiod Forecasting Using Univariate Autoregressions

The **multiperiod regression forecast** h periods into the future based on an AR(p) is computed by estimating the multiperiod regression

$$Y_t = \delta_0 + \delta_1 Y_{t-h} + \cdots + \delta_p Y_{t-p-h+1} + u_t, \quad (14.11)$$

then using the estimated coefficients to compute the forecast h periods in advance.

The **iterated AR forecast** is computed in steps: first compute the one-period ahead forecast, next use that to compute the two-period ahead forecast, and so forth. The two- and three-period ahead iterated forecasts based on an AR(p) are

$$\hat{Y}_{t|t-2} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{t-1|t-2} + \hat{\beta}_2 Y_{t-2} + \hat{\beta}_3 Y_{t-3} + \cdots + \hat{\beta}_p Y_{t-p} \quad (14.12)$$

$$\hat{Y}_{t|t-3} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{t-1|t-3} + \hat{\beta}_2 \hat{Y}_{t-2|t-3} + \hat{\beta}_3 Y_{t-3} + \cdots + \hat{\beta}_p Y_{t-p} \quad (14.13)$$

where the $\hat{\beta}$'s are the OLS estimates of the AR(p) coefficients. Continuing this process ("iterating") produces forecasts further into the future.

The iterated two-quarter ahead forecast is computed by replacing ΔInf_{t-1} in Equation (14.10) with the forecast $\widehat{\Delta Inf}_{t|t-1}$. In Section 12.3, we computed the forecast of $\Delta Inf_{2000:I}$ based on data through 1999:IV using this AR(4) to be $\widehat{\Delta Inf}_{2000:I|1999:IV} = 0.2$. Thus, the two-quarter ahead iterated forecast based on the AR(4) is $\widehat{\Delta Inf}_{2000:II|1999:IV} = 0.02 - 0.21\widehat{\Delta Inf}_{2000:I|1999:IV} - 0.32\Delta Inf_{1999:IV} + 0.19\Delta Inf_{1999:III} - 0.04\Delta Inf_{1999:II} = 0.02 - 0.21 \times 0.2 - 0.32 \times 0.4 + 0.19 \times 0.1 - 0.04 \times 1.1 = -0.2$. According to this iterated AR(4) forecast, based on data through the fourth quarter of 1999, the rate of inflation will fall by 0.2 percentage points between the first and second quarters of 2000.

Both methods for multiperiod univariate forecasting are summarized in Key Concept 14.2.

Multiperiod Forecasting: Multivariate Forecasts

The same two methods for multiperiod forecasting from univariate models can be used in multivariate forecasting regressions.

The multiperiod regression method. In the general multiperiod sion method, all predictors are lagged h periods to produce the h-ahead forecast.

For example, the forecast of ΔInf_t two quarters ahead using four lags ΔInf_t and $Unemp_t$ is computed by first estimating the regression

$$\underset{(0.40)}{\widehat{\Delta Inf}_{t|t-2} = 0.27} - \underset{(0.11)}{0.28\Delta Inf_{t-2}} + \underset{(0.10)}{0.15\Delta Inf_{t-3}} - \underset{(0.11)}{0.21\Delta Inf_{t-4}} - \underset{(0.08)}{0.06\Delta In}$$

$$- \underset{(0.46)}{0.21 Unemp_{t-2}} + \underset{(0.98)}{0.79 Unemp_{t-3}} - \underset{(1.12)}{2.11 Unemp_{t-4}} + \underset{(0.56)}{1.49 Unemp_{t-5}}$$

The two-quarter ahead forecast is computed by substituting the va $\Delta Inf_{1999:I}, \ldots, \Delta Inf_{1999:IV}, Unemp_{1999:I}, \ldots, Unemp_{1999:IV}$ into Equation (14. yields $\widehat{\Delta Inf}_{2000:II|1999:IV} = 0.27 - 0.28\Delta Inf_{1999:IV} + 0.15\Delta Inf_{1999:III} - .21\Delta$ $- 0.06\Delta Inf_{1999:I} - 0.21 Unemp_{1999:IV} + 0.79 Unemp_{1999:III} - 2.11 Unemp$ $1.49 Unemp_{1999:I} = 0.0$.

The three-quarter ahead forecast of ΔInf_t is computed by lagging regressors in Equation (14.14) by one additional quarter, estimating that sion, and computing the forecast, and so forth for forecasts farther into the

The iterated VAR forecast method. The iterated AR method ext a VAR, with the modification that because the VAR has one or mor tional predictors it is necessary to compute intermediate forecasts of predictors.

The two-period ahead **iterated VAR forecast** is computed in two s the first step, the VAR is used to produce one-quarter ahead forecasts o variables in the VAR, as discussed in Section 14.1. In the second step, the casts take the place of the first lagged values in the VAR, that is, the two- ahead forecast is based on the one-period ahead forecast, plus additional specified in the VAR. Repeating this produces the iterated VAR forecast into the future.

As an example, we compute the iterated VAR forecast of $\Delta Inf_{2000:II}$ b data through 1999:IV based on the VAR(4) for ΔInf_t and $Unemp_t$ in Secti (Equations (14.5) and(14.6)). The first step is to compute the one-quarte forecasts $\widehat{\Delta Inf}_{2000:I|1999:IV}$ and $\widehat{Unemp}_{2000:I|1999:IV}$ from that VAR. The $\widehat{\Delta Inf}_{2000:I|1999:IV}$ based on Equation (14.5) was computed in Section 12.3 ar percentage points (Equation (12.18)); a similar calculation based on Equatio shows that $\widehat{Unemp}_{2000:I|1999:IV} = 4.1\%$. In the second step, these forecasts are substituted into Equations (14.5) and (14.6) to produce the two-quarter ahead forecast. Accordingly,

$$\begin{aligned}\widehat{\Delta Inf}_{2000:II|1999:IV} &= 1.32 - 0.36\widehat{\Delta Inf}_{2000:I|1999:IV} - 0.34\Delta Inf_{1999:IV} \\ &\quad + 0.07\Delta Inf_{1999:III} - 0.03\Delta Inf_{1999:II} - 2.68\widehat{Unemp}_{2000:I|1999:IV} \\ &\quad + 3.43 Unemp_{1999:IV} - 1.04 Unemp_{1999:III} + 0.07 Unemp_{1999:II} \\ &= 1.32 - 0.36 \times 0.7 - 0.34 \times 0.4 + 0.07 \times 0.1 - 0.03 \times 1.1 \\ &\quad - 2.68 \times 4.1 + 3.43 \times 4.1 - 1.04 \times 4.2 + 0.07 \times 4.3 \\ &= -0.1.\end{aligned} \quad (14.15)$$

Thus, the iterated VAR(4) forecast, based on data through the fourth quarter of 1999, is that inflation will decline by 0.1 percentage points between the first and second quarters of 2000.

Multiperiod forecasts with multiple predictors are summarized in Key Concept 14.3.

Which Method Should You Use?

Each of the two methods has its advantages and disadvantages. If the autoregressive (or vector autoregressive) model provides a good approximation to the correlations in the data, then the iterated forecast method will tend to produce more precise forecasts. This is because the iterated forecasts use coefficient estimators in a one-period ahead regression, which have a smaller variance (are more efficient) than the estimators from the multiperiod regression.

On the other hand, if the AR or VAR is incorrectly specified and does not provide a good approximation to the correlations in the data, then extrapolating these forecasts by iterating can lead to biased forecasts. Accordingly, if the AR or VAR model is poor, the multiperiod regression forecasts can be more accurate.

Thus there is no easy answer as to whether one method is better than the other. If the difference between the two forecasts is large, this could be an indication that the one-period ahead model is incorrectly specified, and if so the multiperiod ahead forecast tends to be more accurate. Often, however, the differences between the two forecasts is small, as was the case in the inflation forecasts computed in this section, in which case the choice of which method to use can be based on which is most conveniently implemented in your software.

Multiperiod Forecasting with Multiple Predictors

Key Conc 14.3

The multiperiod regression forecast h periods into the future based on p lags each of Y_t and an additional predictor X_t is computed by first estimating the multiperiod regression

$$Y_t = \delta_0 + \delta_1 Y_{t-h} + \cdots + \delta_p Y_{t-p-h+1} + \delta_{p+1} X_{t-h} + \cdots + \delta_{2p} X_{t-p-h+1} + u_t, \quad (14.16)$$

then using the estimated coefficients to make the forecast h periods in advance.

The iterated VAR forecast is computed in steps: first compute the one-period ahead forecasts of all the variables in the VAR, next use those to compute the two-period ahead forecasts, and so forth. The two-period ahead iterated forecast of Y_t based on the two-variable VAR(p) in Key Concept 14.1 is

$$\begin{aligned}\hat{Y}_{t|t-2} &= \hat{\beta}_{10} + \hat{\beta}_{11}\hat{Y}_{t-1|t-2} + \hat{\beta}_{12}Y_{t-2} + \hat{\beta}_{13}Y_{t-3} + \cdots + \hat{\beta}_{1p}Y_{t-p} \\ &\quad + \hat{\gamma}_{11}\hat{X}_{t-1|t-2} + \hat{\gamma}_{12}X_{t-2} + \hat{\gamma}_{13}X_{t-3} + \cdots + \hat{\gamma}_{1p}X_{t-p},\end{aligned} \quad (14.17)$$

where the coefficients in Equation (14.17) are the OLS estimates of the VAR coefficients. Iterating produces forecasts further into the future.

14.3 Orders of Integration and Another Unit Root Test

This section extends the treatment of stochastic trends in Section 12.6 by ad ing two further topics. First, the trends of some time series are not well desc by the random walk model, so we introduce an extension of that model an cuss its implications for regression modeling of such series. Second, we co the discussion of testing for a unit root in time series data and, among other t introduce a second test for a unit root.

Other Models of Trends and Orders of Integration

Recall that the random walk model for a trend, introduced in Section 12.6, specifies that the trend at date t equals the trend at date $t-1$, plus a random error term. If Y_t follows a random walk with drift β_0, then

$$Y_t = \beta_0 + Y_{t-1} + u_t, \quad (14.18)$$

where u_t is serially uncorrelated. Also recall from Section 12.6 that, if a series has a random walk trend, then it has an autoregressive root that equals one.

Although the random walk model of a trend describes the long-run movements of many economic time series, some economic time series have trends that are smoother—that is, vary less from one period to the next—than is implied by Equation (14.18). A different model is needed to describe the trends of such series.

One model of a smooth trend makes the first difference of the trend follow a random walk; that is,

$$\Delta Y_t = \beta_0 + \Delta Y_{t-1} + u_t, \quad (14.19)$$

where u_t is serially uncorrelated. Thus, if Y_t follows Equation (14.19), ΔY_t follows a random walk, so $\Delta Y_t - \Delta Y_{t-1}$ is stationary. The difference of the first differences, $\Delta Y_t - \Delta Y_{t-1}$, is called the **second difference** of Y_t, and is denoted $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$. In this terminology, if Y_t follows Equation (14.19), then its second difference is stationary. If a series has a trend of the form in Equation (14.19), then the first difference of the series has an autoregressive root that equals one.

"Orders of integration" terminology. Some additional terminology is useful for distinguishing between these two models of trends. A series that has a random walk trend is said to be **integrated of order one**, or ***I*(1)**. A series that has a trend of the form in Equation (14.19) is said to be **integrated of order two**, or ***I*(2)**. A series that does not have a stochastic trend and is stationary is said to be **integrated of order zero**, or ***I*(0)**.

The **order of integration** in the $I(1)$ and $I(2)$ terminology is the number of times that the series needs to be differenced for it to be stationary: if Y_t is $I(1)$, then the first difference of Y_t, ΔY_t, is stationary, and if Y_t is $I(2)$, then the second difference of Y_t, $\Delta^2 Y_t$, is stationary. If Y_t is $I(0)$, then Y_t is stationary.

Orders of integration are summarized in Key Concept 14.4.

How to test whether a series is I(2) or I(1). If Y_t is $I(2)$, then ΔY_t is $I(1)$, so that ΔY_t has an autoregressive root that equals one. If, however, Y_t is $I(1)$, then ΔY_t

Orders of Integration, Differencing, and Stationarity

Key Conce[pt] 14.4

- If Y_t is integrated of order 1, that is, if Y_t is $I(1)$, then Y_t has a unit autoregressive root and its first difference, ΔY_t, is stationary.
- If Y_t is integrated of order 2, that is, if Y_t is $I(2)$, then ΔY_t has a unit autoregressive root and its second difference, $\Delta^2 Y_t$, is stationary.
- If Y_t is **integrated of order *d*** (is ***I*(*d*)**), then Y_t must be differenced d times to eliminate its stochastic trend, that is, $\Delta^d Y_t$ is stationary.

is stationary. Thus the null hypothesis that Y_t is $I(2)$ can be tested against the a[lter]native hypothesis that Y_t is $I(1)$ by testing whether ΔY_t has a unit autoregre[ssive] root. If the hypothesis that ΔY_t has a unit autoregressive root is rejected, the[n] hypothesis that Y_t is $I(2)$ is rejected in favor of the alternative that Y_t is $I(1)$.

Examples of I(2) and I(1) series: The price level and the rate of inflat[ion]

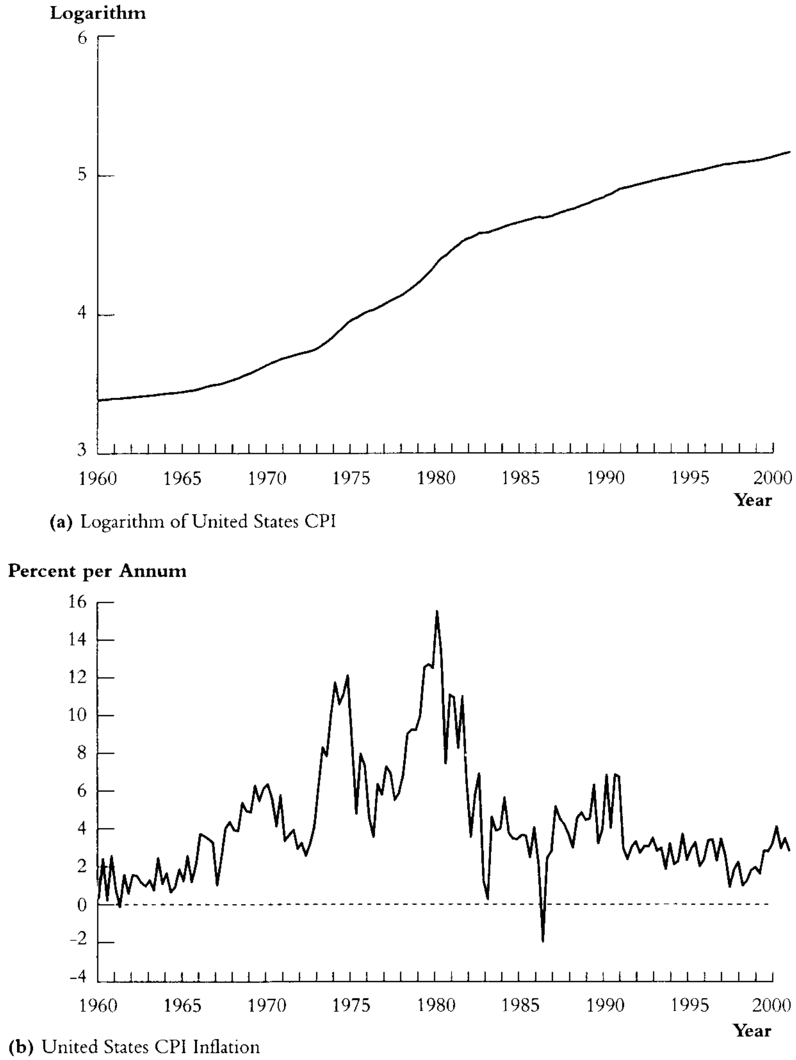
In Chapter 12, we concluded that the rate of inflation in the United States p[os]sibly has a random walk stochastic trend, that is, that the rate of inflation is [...] If inflation is $I(1)$, then its stochastic trend is removed by first differencing, so [...] is stationary. Recall from Section 12.2 (Equation (12.2)) that quarterly infla[tion] at an annual rate is the first difference of the logarithm of the price level, t[imes] 400; that is, $Inf_t = 400\Delta p_t$, where $p_t = \ln(CPI_t)$. Thus treating the rate of infla[tion] as $I(1)$ is equivalent to treating Δp_t as $I(1)$, but this in turn is equivalent to t[reat]ing p_t as $I(2)$. Thus, we have all along been treating the logarithm of the p[rice] level as $I(2)$, even though we have not used that terminology.

The logarithm of the price level, p_t, and the rate of inflation are plotte[d in] Figure 14.1. The long-run trend of the logarithm of the price level (Figure 1[4.1a]) varies more smoothly than the long-run trend in the rate of inflation (Fi[gure] 14.1b). The smoothly varying trend in the logarithm of the price level is ty[pical] of $I(2)$ series.

The DF-GLS Test for a Unit Root

This section continues the discussion of Section 12.6 regarding testing for a [unit] autoregressive root. We first describe another test for a unit autoregressive root[, the] so-called DF-GLS test. Next, in an optional mathematical section, we discuss [why] unit root test statistics do not have normal distributions, even in large sampl[es].

FIGURE 14.1 The Logarithm of the Price Level and the Inflation Rate in the United States, 1960–2000



The trend in the logarithm of prices (Figure 14.1a) is much smoother than the trend in inflation (Figure 14.1b).

The DF-GLS test. The ADF test was the first test developed for testin null hypothesis of a unit root and is the most commonly used test in pra Other tests subsequently have been proposed, however, many of which higher power (Key Concept 3.5) than the ADF test. A test with higher p than the ADF test is more likely to reject the null hypothesis of a unit root ag the stationary alternative when the alternative is true; thus, a more powerfu is better able to distinguish between a unit AR root and a root that is larg less than one.

This section discusses one such test, the so-called **DF-GLS test** devel by Elliott, Rothenberg, and Stock (1996). The test is introduced for the case under the null hypothesis, Y_t has a random walk trend, possibly with drift under the alternative Y_t is stationary around a linear time trend.

The DF-GLS test is computed in two steps. In the first step, the intercep trend are estimated by generalized least squares (GLS; see Section 13.5). The estimation is performed by computing three new variables, V_t, X_{1t}, and X_{2t}, w $V_1 = Y_1$ and $V_t = Y_t - \alpha^* Y_{t-1}$, $t = 2, \ldots, T$, $X_{11} = 1$ and $X_{1t} = 1 - \alpha^*$, $t = 2, \ldots$ and $X_{21} = 1$ and $X_{2t} = t - \alpha^*(t-1)$, where α^* is computed using the formula, $1 - 13.5/T$. Then V_t is regressed against X_{1t} and X_{2t}; that is, OLS is used to mate the coefficients of the population regression equation

$$V_t = \delta_0 X_{1t} + \delta_1 X_{2t} + e_t, \quad (1$$

using the observations $t = 1, \ldots, T$, where e_t is the error term. Note that is no intercept in the regression in Equation (14.20). The OLS estimators $\hat{\delta}$ $\hat{\delta}_1$ are then used to compute a "detrended" version of Y_t, $Y_t^d = Y_t - (\hat{\delta}_0 + \hat{\delta}_1$

In the second step, the Dickey-Fuller test is used to test for a unit autore sive root in Y_t^d, where the Dickey-Fuller regression does not include an inte or a time trend. That is, ΔY_t^d is regressed against Y_{t-1}^d and $\Delta Y_{t-1}^d, \ldots, \Delta Y_{t-p}^d$, w the number of lags p is determined, as usual, either by expert knowledge using a data-based method such as the AIC or BIC as discussed in Section

If the alternative hypothesis is that Y_t is stationary with a mean that mig nonzero but without a time trend, then the preceding steps are modified. Sp ically, α^* is computed using the formula $\alpha^* = 1 - 7/T$, X_{2t} is omitted fro regression in Equation (14.20), and the series Y_t^d is computed as $Y_t^d = Y_t -$

The GLS regression in the first step of the DF-GLS test makes this test complicated than the conventional ADF test, but it is also what improves its ity to discriminate between the null hypothesis of a unit autoregressive and the alternative that Y_t is stationary. This improvement can be substantia example, suppose that Y_t is in fact a stationary AR(1) with autoregressive coefficient $\beta_1 = 0.95$, that there are $T = 200$ observations, and the unit root tests are computed without a time trend (that is, t is excluded from the Dickey-Fuller regression, and X_{2t} is omitted from Equation (14.20)). Then the probability that the ADF test correctly rejects the null hypothesis at the 5% significance level is approximately 31% compared to 75% for the DF-GLS test.

Critical values for DF-GLS test. Because the coefficients on the deterministic terms are estimated differently in the ADF and DF-GLS tests, the tests have different critical values. The critical values for the DF-GLS test are given in Table 14.1. If the DF-GLS test statistic (the t-statistic on Y^d_{t-1} in the regression in the second step) is less than the critical value, then the null hypothesis that Y_t has a unit root is rejected. Like the critical values for the Dickey-Fuller test, the appropriate critical value depends on which version of the test is used, that is, on whether or not a time trend is included (whether or not X_{2t} is included in Equation (14.20)).

Application to Inflation. The DF-GLS statistic, computed for the rate of CPI inflation, Inf_t, over the period 1962:I to 1999:IV, is −1.98 when three lags of ΔY^d_t are included in the Dickey-Fuller regression in the second stage. This value is just less than the 5% critical value in Table 14.1, −1.95, so using the DF-GLS test with three lags leads to rejecting the null hypothesis of a unit root at the 5% significance level. The choice of three lags was based on the AIC (out of a maximum of six lags), which in this case happens to choose the same number of lags as the BIC.

Because the DF-GLS test is better able to discriminate between the unit root null hypothesis and the stationary alternative, one interpretation of this finding is that inflation is in fact stationary, but the Dickey-Fuller test implemented in Section 12.6 failed to detect this (at the 5% level). This conclusion, however, should be tempered by noting that whether the DF-GLS test rejects is, in this application, sensitive to the choice of lag length. If the test is based on four lags, it rejects at the

TABLE 14.1 Critical Values of the DF-GLS Test

Deterministic Regressors (Regressors in Equation (14.20))	10%	5%	1%
Intercept only (X_{1t} only)	−1.62	−1.95	−2.58
Intercept and time trend (X_{1t} and X_{2t})	−2.57	−2.89	−3.48

Source: Fuller (1976) and Elliott, Rothenberg, and Stock (1996, Table 1).

10% but not the 5% level, and if it is instead based on two lags it does not rejec the 10% level. The result is also sensitive to the choice of sample; if the statisti instead computed over the period 1963:I to 1999:IV (that is, dropping just the f year), the test rejects at the 10% level but not at the 5% level. The overall pict therefore is rather ambiguous (as it is based on the ADF test, as discussed follo ing Equation (12.34)) and requires the forecaster to make an informed judgm about whether it is better to model inflation as $I(1)$ or stationary.

Why Do Unit Root Tests Have Nonnormal Distributions?

In Section 12.6, it was stressed that the large-sample normal distribution u which regression analysis relies so heavily does not apply if the regressors are n stationary. Under the null hypothesis that the regression contains a unit root, regressor Y_{t-1} in the Dickey-Fuller regression (and the regressor Y^d_{t-1} in the m ified Dickey-Fuller regression in the second step of the DF-GLS test) is non tionary. The nonnormal distribution of the unit root test statistics is a conseque of this nonstationarity.

To gain some mathematical insight into this nonnormality, consider the si plest possible Dickey-Fuller regression, in which ΔY_t is regressed against the s gle regressor Y_{t-1} and the intercept is excluded. In the notation of Key Conc 12.8, the OLS estimator in this regression is $\hat{\delta} = \sum_{t=1}^{T} Y_{t-1}\Delta Y_t / \sum_{t=1}^{T} Y^2_{t-1}$, so that

$$T\hat{\delta} = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t}{\frac{1}{T^2}\sum_{t=1}^{T} Y^2_{t-1}}. \qquad (14.$$

Consider the numerator in Equation (14.21). Under the additional assu tion that $Y_0 = 0$, a bit of algebra (Exercise 14.5) shows that

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \frac{1}{2}\left[(Y_T/\sqrt{T})^2 - \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2\right]. \qquad (14.$$

Under the null hypothesis, $\Delta Y_t = u_t$, which is serially uncorrelated and a finite variance, so the second term in Equation (14.22) has the probability l

$\frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2 \xrightarrow{p} \sigma_u^2$. Under the assumption that $Y_0 = 0$, the first term in Equation (14.22) can be written $Y_T/\sqrt{T} = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\Delta Y_t = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}u_t$, which in turn obeys the central limit theorem; that is, $Y_T/\sqrt{T} \xrightarrow{d} N(0, \sigma_u^2)$. Thus $(Y_T/\sqrt{T})^2 - \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2 \xrightarrow{d} \sigma_u^2(Z^2 - 1)$, where Z is a standard normal random variable. Recall, however, that the square of a standard normal distribution has a chi-squared distribution with one degree of freedom. It therefore follows from Equation (14.22) that, under the null hypothesis, the numerator in Equation (14.21) has the limiting distribution

$$\frac{1}{T}\sum_{t=1}^{T}Y_{t-1}\Delta Y_t \xrightarrow{d} \frac{\sigma_u^2}{2}(\chi_1^2 - 1). \tag{14.23}$$

The large-sample distribution in Equation (14.23) is different than the usual large-sample normal distribution when the regressor is stationary. Instead, the numerator of the OLS estimator of the coefficient on Y_t in this Dickey-Fuller regression has a distribution that is proportional to a chi-squared distribution with one degree of freedom, minus one.

This discussion has only considered the numerator of $T\hat{\delta}$. The denominator also behaves unusually under the null hypothesis: because Y_t follows a random walk under the null hypothesis, $\frac{1}{T}\sum_{t=1}^{T}Y_{t-1}^2$ does not converge in probability to a constant. Instead, the denominator in Equation (14.21) is a random variable, even in large samples: under the null hypothesis, $\frac{1}{T^2}\sum_{t=1}^{T}Y_{t-1}^2$ converges in distribution jointly with the numerator. The unusual distributions of the numerator and denominator in Equation (14.21) are the source of the nonstandard distribution of the Dickey-Fuller test statistic and the reason that the ADF statistic has its own special table of critical values.

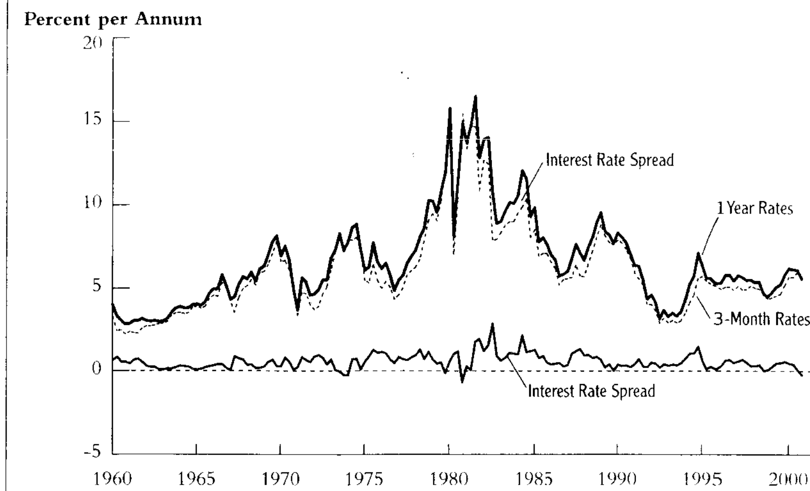
14.4 Cointegration

Sometimes two or more series have the same stochastic trend in common. In this special case, referred to as cointegration, regression analysis can reveal long-run relationships among time series variables, but some new methods are needed.

Cointegration and Error Correction

Two or more time series with stochastic trends can move together so closely over the long run that they appear to have the same trend component, that is, they appear to have a **common trend**. For example, two interest rates on U.S.

FIGURE 14.2 One-Year Interest Rate, Three-Month Interest Rate, and Interest Rate Spre



One-year and three-month interest rates share a common stochastic trend. The spread, or the differe between the two rates does not exhibit a trend. These two interest rates appear to be cointegrated.

government debt are plotted in Figure 14.2. One of the rates is the interest on 90-day U.S. Treasury bills, at an annual rate ($R90_t$); the other is the int rate on a one-year U.S. Treasury bond ($R1yr_t$); these interest rates are disc in Appendix 14.1. The interest rates exhibit the same long-run tendenci trends: both were low in the 1960s, both rose through the 1970s to peaks i early 1980s, then both fell through the 1990s. Moreover, the difference bet the two series, $R1yr_t - R90_t$, which is called the "spread" between the two i est rates and is also plotted in Figure 14.2, does not appear to have a trend. is, subtracting the 90-day interest rate from the one-year interest rate appea eliminate the trends in both of the individual rates. Said differently, althoug two interest rates differ, they appear to share a common stochastic trend: be the trend in each individual series is eliminated by subtracting one series fro other, the two series must have the same trend, that is, they must have a com stochastic trend.

Two or more series that have a common stochastic trend are said to be **cointegrated**. The formal definition of cointegration (due to Granger, 1983) is given in Key Concept 14.5. In this section, we introduce a test for whether cointegration is present, discuss estimation of the coefficients of regressions relating cointegrated variables, and illustrate the use of the cointegrating relationship for forecasting. The discussion initially focuses on the case that there are only two variables, X_t and Y_t.

Vector error correction model. Until now, we have eliminated the stochastic trend in an $I(1)$ variable Y_t by computing its first difference, ΔY_t; the problems created by stochastic trends were then avoided by using ΔY_t instead of Y_t in time series regressions. If X_t and Y_t are cointegrated, however, another way to eliminate the trend is to compute the difference $Y_t - \theta X_t$. Because the term $Y_t - \theta X_t$ is stationary, it too can be used in regression analysis.

In fact, if X_t and Y_t are cointegrated, the first differences of X_t and Y_t can be modeled using a VAR, augmented by including $Y_{t-1} - \theta X_{t-1}$ as an additional regressor:

$$\Delta Y_t = \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \cdots + \gamma_{1p}\Delta X_{t-p} + \alpha_1(Y_{t-1} - \theta X_{t-1}) + u_{1t} \qquad (14.24)$$

$$\Delta X_t = \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \cdots + \gamma_{2p}\Delta X_{t-p} + \alpha_2(Y_{t-1} - \theta X_{t-1}) + u_{2t}. \qquad (14.25)$$

The term $Y_t - \theta X_t$ is called the **error correction term**. The combined model in Equations (14.24) and (14.25) is called a **vector error correction model** (VECM). In a VECM, past values of $Y_t - \theta X_t$ help to predict future values of ΔY_t and/or ΔX_t.

How Can You Tell Whether Two Variables Are Cointegrated?

There are three ways to decide whether two variables can plausibly be modeled as cointegrated: use expert knowledge and economic theory, graph the series and see whether they appear to have a common stochastic trend, and perform statistical tests for cointegration. All three methods should be used in practice.

First, you must use your expert knowledge of these variables to decide whether cointegration is in fact plausible. For example, the two interest rates in Figure 14.2 are linked together by the so-called expectations theory of the term structure of interest rates. According to this theory, the interest rate on January 1

Cointegration

Key Concep 14.5

Suppose X_t and Y_t are integrated of order one. If, for some coefficient θ, $Y_t - \theta X_t$ is integrated of order zero, then X_t and Y_t are said to be **cointegrated**. The coefficient θ is called the **cointegrating coefficient**.

If X_t and Y_t are cointegrated, then they have the same, or common, stochastic trend. Computing the difference $Y_t - \theta X_t$ eliminates this common stochastic trend.

on the one-year Treasury bond is the average of the interest rate on a 90-day Tr sury bill for the first quarter of the year and the expected interest rates on fut 90-day Treasury bills issued in the second, third, and fourth quarters of the ye if not, then investors could expect to make money by holding either the one-y Treasury note or a sequence of four 90-day Treasury bills, and they would bid prices until the expected returns are equalized. If the 90-day interest rate ha random walk stochastic trend, this theory implies that this stochastic trend is inh ited by the one-year interest rate and that the difference between the two ra that is, the spread, is stationary. Thus, the expectations theory of the term str ture implies that, if the interest rates are $I(1)$, then they will be cointegrated w a cointegrating coefficient of $\theta = 1$ (Exercise 14.2).

Second, visual inspection of the series helps to identify cases in which co tegration is plausible. For example, the graph of the two interest rates in Fig 14.2 shows that each of the series appears to be $I(1)$ but that the spread appe to be $I(0)$, so that the two series appear to be cointegrated.

Third, the unit root testing procedures introduced so far can be extendec test for cointegration. The insight on which these tests are based is that if Y_t X_t are cointegrated with cointegrating coefficient θ, then $Y_t - \theta X_t$ is statione otherwise, $Y_t - \theta X_t$ is nonstationary (is $I(1)$). The hypothesis that Y_t and X_t are cointegrated (that is, that $Y_t - \theta X_t$ is $I(1)$) therefore can be tested by testing null hypothesis that $Y_t - \theta X_t$ has a unit root; if this hypothesis is rejected, ther and X_t can be modeled as cointegrated. The details of this test depend on whet the cointegrating coefficient θ is known.

Testing for cointegration when θ is known. In some cases expert knowle or economic theory suggests values of θ. When θ is known, the Dickey-Fuller

DF-GLS unit root tests can be used to test for cointegration by first constructing the series $z_t = Y_t - \theta X_t$, then testing the null hypothesis that z_t has a unit autoregressive root.

Testing for cointegration when θ is unknown. If the cointegrating coefficient θ is unknown then it must be estimated prior to testing for a unit root in the error correction term. This preliminary step makes it necessary to use different critical values for the subsequent unit root test.

Specifically, in the first step the cointegrating coefficient θ is estimated by OLS estimation of the regression

$$Y_t = \alpha + \theta X_t + z_t. \tag{14.26}$$

In the second step, a Dickey-Fuller t-test (with an intercept but no time trend) is used to test for a unit root in the residual from this regression, $\hat{z}_t$. This two-step procedure is called the Engle-Granger Augmented Dickey-Fuller test for cointegration, or **EG-ADF** (Engle and Granger, 1987).

Critical values of the EG-ADF statistic are given in Table 14.2.[2] The critical values in the first row apply when there is a single regressor in Equation (14.26), so that there are two cointegrated variables (X_t and Y_t). The subsequent rows apply to the case of multiple cointegrated variables, which is discussed at the end of this section.

Estimation of Cointegrating Coefficients

If X_t and Y_t are cointegrated, then the OLS estimator of the coefficient in the cointegrating regression in Equation (14.26) is consistent. However, in general the OLS estimator has a nonnormal distribution, and inferences based on its t-statistics can be misleading whether or not those t-statistics are computed using HAC standard errors. Because of these drawbacks of the OLS estimator of θ, econometricians have developed a number of other estimators of the cointegrating coefficient.

One such estimator of θ that is simple to use in practice is the so-called **dynamic OLS (DOLS)** estimator (Stock and Watson, 1993). The DOLS esti-

[2]The critical values in Table 14.2 are taken from Fuller (1976) and Phillips and Ouliaris (1990). Following a suggestion by Hansen (1992), the critical values in Table 14.2 are chosen so that they apply whether or not X_t and Y_t have drift components.

TABLE 14.2 Critical Values for the Engle-Granger ADF statistic

Number of X's in Equation (14.26)	10%	5%
1	−3.12	−3.41
2	−3.52	−3.80
3	−3.84	−4.16
4	−4.20	−4.49

mator is based on a modified version of Equation (14.26) that includes pas sent, and future values of the change in X_t:

$$Y_t = \beta_0 + \theta X_t + \sum_{j=-p}^{p} \delta_j \Delta X_{t-j} + u_t. \quad ($$

Thus, in Equation (14.27), the regressors are $X_t, \Delta X_{t+p}, \ldots, \Delta X_{t-p}$. The I estimator of θ is the OLS estimator of θ in the regression of Equation (14.

If X_t and Y_t are cointegrated, then the DOLS estimator is efficient ir samples. Moreover, statistical inferences about θ and the δ's in Equation (based on HAC standard errors are valid. For example, the t-statistic const using the DOLS estimator with HAC standard errors has a standard norm tribution in large samples.

One way to interpret Equation (14.27) is to recall from Section 13. cumulative dynamic multipliers can be computed by modifying the distr lag regression of Y_t on X_t and its lags. Specifically, in Equation (13.7), the c lative dynamic multipliers were computed by regressing Y_t on ΔX_t, lags o and X_{t-r}; the coefficient on X_{t-r} in that specification is the long-run cum dynamic multiplier. Similarly, if X_t were strictly exogenous, then in Eq (14.27), the coefficient on X_t, θ, would be the long-run cumulative mult that is, the long-run effect on Y of a change in X. If X_t is not strictly exog then the coefficients do not have this interpretation. Nevertheless, because Y_t have a common stochastic trend if they are cointegrated, the DOLS esti is consistent even if X_t is endogenous.

The DOLS estimator is not the only efficient estimator of the cointeg coefficient. The first such estimator was developed by Søren Johansen (Joh 1988). For a discussion of Johansen's method and of other ways to estima cointegrating coefficient, see Hamilton (1994, Chapter 20).

Even if economic theory does not suggest a specific value of the cointegrating coefficient, it is important to check whether the estimated cointegrating relationship makes sense in practice. Because cointegration tests can be misleading (they can improperly reject the null hypothesis of no cointegration more frequently than they should, and frequently they improperly fail to reject the null), it is especially important to rely on economic theory, institutional knowledge, and common sense when estimating and using cointegrating relationships.

Extension to Multiple Cointegrated Variables

The concepts, tests, and estimators discussed here extend to more than two variables. For example, if there are three variables, Y_t, X_{1t}, and X_{2t}, each of which is $I(1)$, then they are cointegrated with cointegrating coefficients θ_1 and θ_2 if $Y_t - \theta_1 X_{1t} - \theta_2 X_{2t}$ is stationary. When there are three or more variables, there can be multiple cointegrating relationships. For example, consider modeling the relationship among three interest rates: the three-month rate, the one-year rate, and the five-year rate ($R5yr$). If they are $I(1)$, then the expectations theory of the term structure of interest rates suggests that they will all be cointegrated. One cointegrating relationship suggested by the theory is $R90_t - R1yr_t$, and a second relationship is $R90_t - R5yr_t$. (The relationship $R1yr_t - R5yr_t$ is also a cointegrating relationship, but it contains no additional information beyond that in the other relationships because it is perfectly multicollinear with the other two cointegrating relationships.)

The EG-ADF procedure for testing for a single cointegrating relationship among mulitple variables is the same as for the case of two variables, except that the regression in Equation (14.26) is modified so that both X_{1t} and X_{2t} are regressors; the critical values for the EG-ADF test are given in Table 14.2, where the appropriate row depends on the number of regressors in the first-stage OLS cointegrating regression. The DOLS estimator of a single cointegrating relationship among multiple X's involves including the level of each X along with leads and lags of the first difference of each X. Tests for multiple cointegrating relationships can be performed using the system methods, such as Johansen's (1988) method, and the DOLS estimator can be extended to multiple cointegrating relationships by estimating multiple equations, one for each cointegrating relationship. For additional discussion of cointegration methods for multiple variables, see Hamilton (1994).

A cautionary note. If two or more variables are cointegrated then the error correction term can help to forecast these variables and, possibly, other related variables. However, cointegration requires the variables to have the same stochastic trends. Trends in economic variables typically arise from complex interactions of

disparate forces, and closely related series can have different trends for subtle re
sons. If variables that are not cointegrated are incorrectly modeled using a VEC
then the error correction term will be $I(1)$; this introduces a trend into the fo
cast that can result in poor out-of-sample forecast performance. Thus forecasti
using a VECM must be based on a combination of compelling theoretical arg
ments in favor of cointegration and careful empirical analysis.

Application to Interest Rates

As discussed above, the expectations theory of the term structure of interest ra
implies that, if two interest rates of different maturities are $I(1)$, then they will
cointegrated with a cointegrating coefficient of $\theta = 1$, that is, the spread between t
two rates will be stationary. Inspection of Figure 14.2 provides qualitative supp
for the hypothesis that the one-year and three-month interest rates are coint
grated. We first use unit root and cointegration test statistics to provide mo
formal evidence on this hypothesis, then estimate a vector error correction mod
for these two interest rates.

Unit root and cointegration tests. Various unit root and cointegration te
statistics for these two series are reported in Table 14.3. The unit root test stati
tics in the first two rows examine the hypothesis that the two interest rates, t
three-month rate ($R90$) and the one-year rate ($R1yr$), individually have a unit roo
Two of the four statistics in the first two rows fail to reject this hypothesis at t
10% level, and three of the four fail to reject at the 5% level. The exception is t
ADF statistic evaluated for the 90-day Treasury bill rate (−2.96), which rejects t
unit root hypothesis at the 5% level. The ADF and DF-GLS statistics lead

TABLE 14.3 Unit Root and Cointegration Test Statistics for Two Interest Rates

Series	ADF Statistic	DF-GLS Statistic
$R90$	−2.96*	−1.88⁺
$R1yr$	−2.22	−1.37
$R1yr - R90$	−6.31**	−5.59**
$R1yr - 1.046R90$	−6.97**	—

$R90$ is the interest rate on 90-day U.S. Treasury bills, at an annual rate, and $R1yr$ is the interest rate
one-year U.S. Treasury bonds. Regressions were estimated using quarterly data over the peri
1962:I–1999:IV. The number of lags in the unit root test statistic regressions were chosen by AIC (
lags maximum). Unit root test statistics are significant at the ⁺10%, *5%, or **1% significance level.

different conclusions for this variable (the ADF test rejects the unit root hypothesis at the 5% level while the DF-GLS test does not), which means that we must exercise some judgment in deciding whether these variables are plausibly modeled as $I(1)$. Taken together, these results suggest that the interest rates are plausibly modeled as $I(1)$.

The unit root statistics for the spread, $R1yr_t - R90_t$, test the further hypothesis that these variables are not cointegrated against the alternative that they are. The null hypothesis that the spread contains a unit root is rejected at the 1% level using both unit root tests. Thus we reject the hypothesis that the series are not cointegrated against the alternative that they are, with a cointegrating coefficient $\theta = 1$. Taken together, the evidence in the first three rows of Table 14.3 suggests that these variables plausibly can be modeled as cointegrated with $\theta = 1$.

Because in this application economic theory suggests a value for θ (the expectations theory of the term structure suggests that $\theta = 1$) and because the error correction term is $I(0)$ when this value is imposed (the spread is stationary), in principle it is not necessary to use the EG-ADF test, in which θ is estimated. Nevertheless, we compute the test as an illustration. The first step in the EG-ADF test is to estimate θ by the OLS regression of one variable on the other; the result is

$$\widehat{R1yr_t} = 0.361 + 1.046R90_t,\ \overline{R}^2 = 0.973. \quad (14.28)$$

The second step is computing the ADF statistic for the residual from this regression, $\hat{z}_t$. The result, given in the final row of Table 14.3, is less than the 1% critical value of −3.96 in Table 14.2, so the null hypothesis that $\hat{z}_t$ has a unit autoregressive root is rejected. This statistic also points towards treating the two interest rates as cointegrated. Note that no standard errors are presented in Equation (14.28) because, as previously discussed, the OLS estimator of the cointegrating coefficient has a nonnormal distribution and its t-statistic is not normally distributed, so presenting standard errors (HAC or otherwise) would be misleading.

A vector error correction model of the two interest rates. If Y_t and X_t are cointegrated, then forecasts of ΔY_t and ΔX_t can be improved by augmenting a VAR of ΔY_t and ΔX_t by the lagged value of the error correction term, that is, by computing forecasts using the VECM in Equations (14.24) and(14.25). If θ is known, then the unknown coefficients of the VECM can be estimated by OLS, including $z_{t-1} = Y_{t-1} - \theta X_{t-1}$ as an additional regressor. If θ is unknown, then the

VECM can be estimated using $\hat{z}_{t-1}$ as a regressor, where $\hat{z}_t = Y_t - \hat{\theta}X_t$, whe an estimator of θ.

In the application to the two interest rates, theory suggests that $\theta = 1$, ar unit root tests supported modeling the two interest rates as cointegrated v cointegrating coefficient of 1. We therefore specify the VECM using the retically suggested value of $\theta = 1$, that is, by adding the lagged value of the sp $R1yr_{t-1} - R90_{t-1}$ to a VAR in $\Delta R1yr_t$ and $\Delta R90_t$. Specified with two lags o differences, the resulting VECM is

$$\begin{aligned}\widehat{\Delta R90_t} &= \underset{(0.17)}{0.14} - \underset{(0.32)}{0.24}\Delta R90_{t-1} - \underset{(0.34)}{0.44}\Delta R90_{t-2} - \underset{(0.39)}{0.01}\Delta R1yr_{t-1}\\ &\quad + \underset{(0.27)}{0.15}\Delta R1yr_{t-2} - \underset{(0.27)}{0.18}(R1yr_{t-1} - R90_{t-1}) \quad (1\end{aligned}$$

$$\begin{aligned}\widehat{\Delta R1yr_t} &= \underset{(0.16)}{0.36} - \underset{(0.30)}{0.14}\Delta R90_{t-1} - \underset{(0.29)}{0.33}\Delta R90_{t-2} - \underset{(0.35)}{0.11}\Delta R1yr_{t-1}\\ &\quad + \underset{(0.25)}{0.10}\Delta R1yr_{t-2} - \underset{(0.24)}{0.52}(R1yr_{t-1} - R90_{t-1}) \quad (1\end{aligned}$$

In the first equation, none of the coefficients are individually significant 5% level and the coefficients on the lagged first differences of the interest rat not jointly significant at the 5% level. In the second equation, the coefficier the lagged first differences are not jointly significant, but the coefficient o lagged spread (the error correction term), which is estimated to be −.52, t-statistic of −2.17, so it is statistically significant at the 5% level. Although la values of the first difference of the interest rates are not useful for predicting f interest rates, the lagged spread does help to predict the change in the one Treasury bond rate. When the one-year rate exceeds the 90-day rate, the year rate is forecasted to fall in the future.

14.5 Conditional Heteroskedasticity

The phenomenon that some times are tranquil while others are not—that is volatility comes in clusters—shows up in many economic time series. Thi tion presents a pair of models for quantifying **volatility clustering** or, as it i known, **conditional heteroskedasticity**.

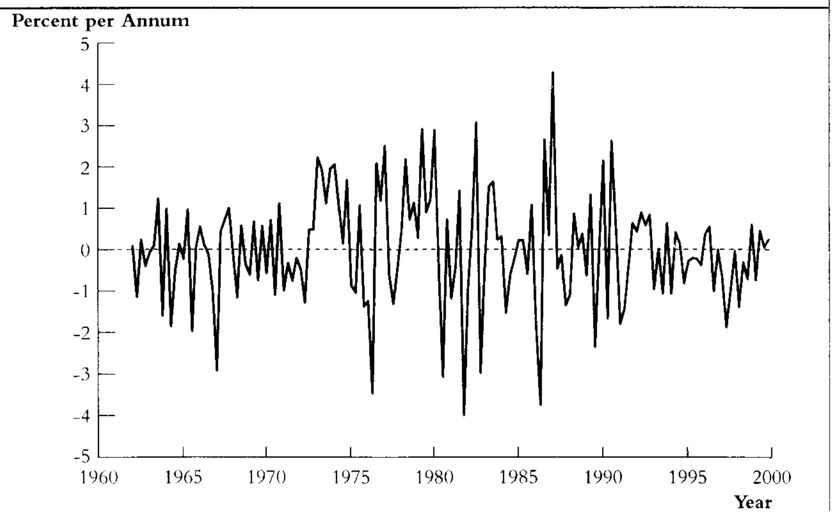
Volatility Clustering

Section 12.7 had a curious empirical result: the root mean squared forecast error of the pseudo out-of-sample forecasts of inflation from 1996 to 1999, produced using the four-lag Phillips curve, was 0.75 percentage points, whereas the standard error of the OLS regression that produced those forecasts was 1.47. That is, the out-of-sample errors were half the size of the in-sample errors! A forecaster faced with this happy situation might be forgiven for crowing about this to his or her clients. Might it be, however, that forecasting is simply easier at some times than at others, and the late 1990s was just one of those easy times?

Visual inspection of the residuals from the four-lag Phillips curve (Equation (14.5)), plotted in Figure 14.3, suggest so: these residuals exhibit volatility clustering. In the late 1970s and early 1980s the absolute forecast errors often exceeded two percentage points. In the 1960s and 1990s, however, the absolute forecast errors typically are less than one percentage point.

Volatility clustering is evident in many financial time series. An example discussed in Section 12.2 is shown in Figure 12.2d, a plot of 1,771 daily returns on the NYSE Composite Index of stock prices from 1990 to 1998. The absolute

FIGURE 14.3 Residuals from Phillips Curve in Equation (14.5)



The residuals from the Phillips curve show volatility clustering. Variability is relatively low in the 1960s and 1990s and higher in the 1970s and 1980s.

daily percentage changes were, on average, larger in 1991 and 1998 than in and 1995. Within any given year, some months have greater volatility than ers. Like the Phillips curve residuals, these percentage price changes have exte periods of high volatility and extended periods of relative tranquility.

Volatility clustering can be thought of as clustering of the variance of the term over time: if the regression error has a small variance in one period, its ance tends to be small in the next period too. In other words, volatility clu ing implies that the error exhibits time-varying heteroskedasticity.

Autoregressive Conditional Heteroskedasticity

Two models of volatility clustering are the autoregressive conditional eroskedasticity (ARCH) model and its extension, the generalized AI (GARCH) model.

ARCH. Consider the ADL(1,1) regression

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t. \quad (1$$

In the **ARCH** model, developed by the econometrician Robert Engle (E 1982), the error u_t is modeled as being normally distributed with mean zero variance σ_t^2, where σ_t^2 depends on past squared values u_t. Specifically, the AR model of order p, denoted ARCH(p), is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2, \quad (1$$

where $\alpha_0, \alpha_1, \ldots, \alpha_p$ are unknown coefficients. If these coefficients are pos then if recent squared errors are large the ARCH model predicts that the cu squared error will be large in magnitude in the sense that its variance, σ_t^2, is la

Although it is described here for the ADL(1,1) model in Equation (14.31) ARCH model can be applied to the error variance of any time series regre model with an error that has a conditional mean of zero, including higher ADL models, autoregressions, and time series regressions with multiple predic

GARCH. The generalized ARCH (**GARCH**) model, developed by the ec metrician Timothy Bollerslev (1986), extends the ARCH model to let σ_t^2 de on its own lags as well as lags of the squared error. The GARCH(p,q) model

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \phi_1 \sigma_{t-1}^2 \cdots + \phi_q \sigma_{t-q}^2, \quad (1$$

where $\alpha_0, \alpha_1, \ldots, \alpha_p, \phi_1, \ldots, \phi_q$ are unknown coefficients.

The ARCH model is analogous to a distributed lag model, and the GARCH model is analogous to an ADL model. As discussed in Appendix 13.2, the ADL model (when appropriate) can provide a more parsimonious model of dynamic multipliers than the distributed lag model. Similarly, by incorporating lags of σ_t^2, the GARCH model can capture slowly changing variances with fewer parameters than the ARCH model.

An important application of ARCH and GARCH models is to measuring and forecasting the time-varying volatility of returns on financial assets, particularly assets observed at high sampling frequencies such as the daily stock returns in Figure 12.2d. In such applications the return itself is often modeled as unpredictable, so the regression in Equation (14.31) only includes the intercept.

Estimation and inference. ARCH and GARCH models are estimated by the method of maximum likelihood (Appendix 9.2). The estimators of the ARCH and GARCH coefficients are normally distributed in large samples, so in large samples t-statistics have standard normal distributions and confidence intervals for a coefficient can be constructed as its maximum likelihood estimate ±1.96 standard errors.

Application to Inflation Forecasts

The four-lag Phillips curve, estimated by OLS in Equation (14.5), was re-estimated using a GARCH(1,1) model for the error term over the same period, yielding

$$\begin{aligned}\widehat{\Delta Inf_t} &= \underset{(0.33)}{1.29} - \underset{(0.10)}{0.41}\Delta Inf_{t-1} - \underset{(0.09)}{0.31}\Delta Inf_{t-2} + \underset{(0.11)}{0.02}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4} \\ &\quad - \underset{(0.34)}{2.50}Unemp_{t-1} + \underset{(0.71)}{2.76}Unemp_{t-2} + \underset{(0.81)}{0.15}Unemp_{t-3} - \underset{(0.40)}{0.64}Unemp_{t-4};\end{aligned} \tag{14.34}$$

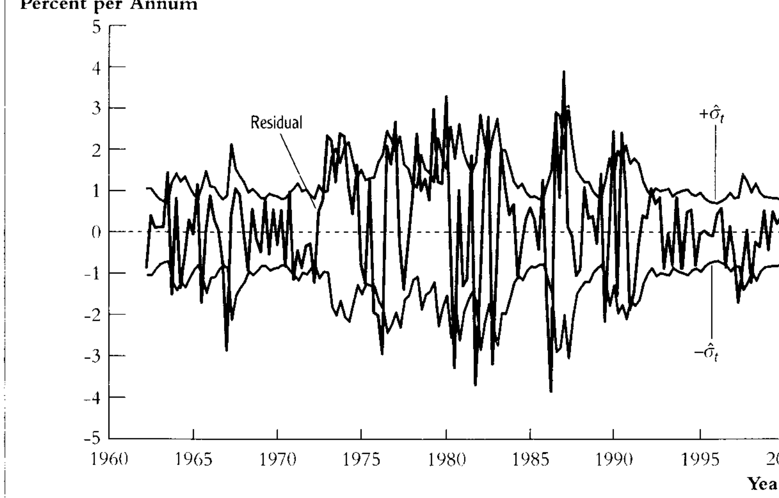
$$\hat{\sigma}_t^2 = \underset{(0.14)}{0.26} + \underset{(0.20)}{0.47}u_{t-1}^2 + \underset{(0.17)}{0.45}\sigma_{t-1}^2. \tag{14.35}$$

The two coefficients in the GARCH model (the coefficients on u_{t-1}^2 and σ_{t-1}^2) are both individually statistically significant at the 5% significance level, and the joint hypothesis that both coefficients are zero also can be rejected at the 5% significance level. Thus, we can reject the null hypothesis that the Phillips curve errors are homoskedastic against the alternative that they are conditionally heteroskedastic.

The ADL coefficients estimated by OLS (Equation (14.5)) and by max likelihood with the GARCH model (Equation (14.34)) are slightly different. two GARCH coefficients in Equation (14.35) were exactly zero, then the tw of estimates would be identical. However, these coefficients are nonzero; be maximum likelihood estimation estimates the coefficients in Equations (14.3 (14.35) simultaneously, the two sets of estimated ADL coefficients differ.

The predicted variances, $\hat{\sigma}_t^2$, can be computed using the coefficients in tion (14.35) and the residuals from Equation (14.34). These residuals are p in Figure 14.4, along with bands of plus or minus one predicted standard tion (that is, $\pm\hat{\sigma}_t$) based on the GARCH(1,1) model. These bands quanti changing volatility of the Phillips curve residuals over time. During the 1980s, these conditional standard deviation bands are wide, indicating co erable volatility in the Phillips curve regression error and thus conside

FIGURE 14.4 Residuals from the Phillips Curve in Equation (14.34) and GARCH(1,1) Ba



The GARCH(1,1) bands, which are $\pm\hat{\sigma}_t$ computed using Equation (14.35), are narrow when the conditional variance is small and wide when it is large. The forecast interval is narrower at the beginning and end of the sample when $\hat{\sigma}_t$ is small.

uncertainty about the resulting inflation forecasts. In the late 1960s and late 1990s, however, these bands are tight.

With these conditional standard deviation bands in hand, we now can return to the question with which we started this section: was the late 1990s an unusually tranquil period for forecasting inflation? The estimated conditional variances suggest that it was. For example, the predicted standard deviation in 1993:IV is $\hat{\sigma}_{1993:IV} = 0.97$, well less than the OLS standard error of the regression in Equation (14.5), which was 1.47. The actual pseudo out-of-sample RMSFE of 0.75 is still less than the GARCH estimate of 0.97, but not by much.

14.6 Conclusion

This part of the book has covered some of the most frequently used tools and concepts of time series regression. Many other tools for analyzing economic time series have been developed for specific applications. If you are interested in learning more about economic forecasting, see the introductory textbooks by Enders (1995) and Diebold (2000). For an advanced, modern, and comprehensive treatment of econometrics with time series data, see Hamilton (1994).

Summary

1. Vector autoregressions model a "vector" of k time series variables as each depends on its own lags and the lags of the $k - 1$ other series. The forecasts of each of the time series produced by a VAR are mutually consistent, in the sense that they are based on the same information.
2. Forecasts two or more periods ahead can be computed either by iterating forward a one-step ahead model (an AR or a VAR) or by estimating a multiperiod ahead regression.
3. Two series that share a common stochastic trend are cointegrated; that is, Y_t and X_t are cointegrated if Y_t and X_t are $I(1)$ but $Y_t - \theta X_t$ is $I(0)$. If Y_t and X_t are cointegrated, the error correction term $Y_t - \theta X_t$ can help to predict ΔY_t and/or ΔX_t. A vector error correction model is a VAR model of ΔY_t and ΔX_t, augmented to include the lagged error correction term.
4. Volatility clustering—when the variance of a series is high in some periods and low in others—is common in economic time series, especially financial time series.

5. The ARCH model of volatility clustering expresses the conditional vari[...] the regression error as a function of recent squared regression error[...] GARCH model augments the ARCH model to include lagged condition[...] ances as well. Estimated ARCH and GARCH models produce forecast i[...] with widths that depend on the volatility of the most recent regression re[...]

Key Terms

vector autoregression (VAR) (534)
multiperiod regression forecast (542)
iterated AR forecast (542)
iterated VAR forecast (543)
second difference (546)
$I(0)$, $I(1)$, and $I(2)$ (546)
order of integration (546)
integrated of order d ($I(d)$) (547)
DF-GLS test (549)
common trend (553)
error correction term (554)
vector error correction model (554)
cointegration (555)
cointegrating coefficient (555)
EG-ADF test (556)
DOLS estimator (557)
volatility clustering (561)
conditional heteroskedasticity (561)
ARCH (563)
GARCH (563)

Review the Concepts

14.1 A macroeconomist wants to construct forecasts for the following mac[...] nomic variables: GDP, consumption, investment, government pur[...] exports, imports, short-term interest rates, long-term interest rates, a[...] rate of price inflation. He has quarterly time series for each of these va[...] from 1970–2001. Should he estimate a VAR for these variables and u[...] for forecasting? Why or why not? Can you suggest an alternative app[...]

14.2 Suppose that Y_t follows a stationary AR(1) model with $\beta_0 = 0$ and β_1[...] If $Y_t = 5$, what is your forecast of Y_{t+2} (that is, what is $Y_{t+2|t}$)? What i[...] for $h = 30$? Does this forecast for $h = 30$ seem reasonable to you?

14.3 A version of the permanent income theory of consumption implies t[...] logarithm of real GDP (Y) and the logarithm of real consumption ([...] cointegrated with a cointegrating coefficient equal to 1. Explain ho[...]

would investigate this implication by (a) plotting the data, and (b) using a statistical test.

14.4 Consider the ARCH model, $\sigma_t^2 = 1.0 + 0.8u_{t-1}^2$. Explain why this will lead to volatility clustering. (*Hint:* What happens when u_{t-1}^2 is unusually large?)

14.5 The DF-GLS test for a unit root has higher power than the Dickey-Fuller test. Why should you use a more powerful test?

Exercises

14.1 Suppose that Y_t follows a stationary AR(1) model, $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$.

***a.** Show that the h-period ahead forecast of Y_t is given by $Y_{t+h|t} = \mu_Y + \beta_1^h(Y_t - \mu_Y)$, where $\mu_Y = \beta_0/(1 - \beta_1)$.

b. Suppose that X_t is related to Y_t by $X_t = \sum_{i=0}^{\infty}\delta^i Y_{t+i|t}$, where $|\delta| < 1$. Show that $X_t = \frac{\mu_Y}{1-\delta} + \frac{Y_t - \mu_Y}{1-\beta_1\delta}$.

14.2 One version of the expectations theory of the term structure of interest rates holds that a long-term rate equals the average of the expected values of short-term interest rates into the future, plus a term premium that is $I(0)$. Specifically, let Rk_t denote a k-period interest rate, let $R1_t$ denote a one-period interest rate, and let e_t denote an $I(0)$ term premium. Then $Rk_t = \frac{1}{k}\sum_{i=1}^{k} R1_{t+i|t} + e_t$, where $R1_{t+i|t}$ is the forecast made at date t of the value of $R1$ at date $t + i$. Suppose that $R1_t$ follows a random walk, so that $R1_t = R1_{t-1} + u_t$.

a. Show that $Rk_t = R1_t + e_t$.

b. Show that Rk_t and $R1_t$ are cointegrated. What is the cointegrating coefficient?

c. Now suppose that $\Delta R1_t = 0.5\Delta R1_{t-1} + u_t$. How does your answer to (b) change?

d. Now suppose that $R1_t = 0.5R1_{t-1} + u_t$. How does your answer to (b) change?

14.3 Suppose that u_t follows the ARCH process, $\sigma_t^2 = 1.0 + 0.5u_{t-1}^2$.

***a.** Let $E(u_t^2) = \text{var}(u_t)$ be the unconditional variance of u_t. Show that $\text{var}(u_t) = 2$.

b. Suppose that the distribution of u_t conditional on lagged values o $N(0, \sigma_t^2)$. If $u_{t-1} = 0.2$, what is $\Pr(-3 \le u_t \le 3)$? If $u_{t-1} = 2.0$, what $\Pr(-3 \le u_t \le 3)$?

14.4 Suppose that Y_t follows the AR(p) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p$ u_t, where $E(u_t | Y_{t-1}, Y_{t-2}, \cdots) = 0$. Let $Y_{t+h|t} = E(Y_{t+h} | Y_t, Y_{t-1}, \ldots)$. S that $Y_{t+h|t} = \beta_0 + \beta_1 Y_{t-1+h|t} + \cdots + \beta_p Y_{t-p+h|t}$ for $h > p$.

14.5 Verify Equation (14.22). (*Hint:* use $\sum_{t=1}^{T} Y_t^2 = \sum_{t=1}^{T}(Y_{t-1} + \Delta Y_t)^2$ to show $\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 = 2\sum_{t=1}^{T} Y_{t-1}\Delta Y_t + \sum_{t=1}^{T}\Delta Y_t^2$ and solve for $\sum_{t=1}^{T} Y_{t-1}\Delta Y_t$.)

APPENDIX 14.1 U.S. Financial Data Used in Chapter 14

The interest rates on three-month U.S. Treasury bills and on one-year U.S. Treasury b are the monthly average of their daily rates, converted to an annual basis, as reporte the U.S. Federal Reserve Bank. The quarterly data used in this chapter are the mo average interest rates for the final month in the quarter.