Chapter 9

Regression With Time-Series Data: Stationary Variables

Learning Objectives

Based on the material in this Chapter, you should be able to:

1. Explain why lags are important in models that use time-series data, and the ways in which lags can be included in dynamic econometric models.
2. Explain what is meant by a serially correlated time series, and how we measure serial correlation.
3. Specify, estimate, and interpret the estimates from a finite distribute lag model.
4. Explain the nature of regressions that involve lagged variables and the number of observations that are available.
5. Specify and explain how the multiple regression assumptions are modified to accommodate time series data.
6. Compute the autocorrelations for a time-series, graph the corresponding correlogram, and use it to test for serial correlation.
7. Use a correlogram of residuals to test for serially correlated errors.
8. Use a Lagrange multiplier test for serially correlated errors.
9. Compute HAC standard errors for least squares estimates. Explain why they are used.
10. Describe the properties of an AR(1) error.
11. Compute nonlinear least squares estimates for a model with an AR(1) error.
12. Test whether an ARDL(1, 1) model can be written as an AR(1) error model.
13. Specify and estimate autoregressive distributed lag models. Use serial correlation checks, significance of coefficients and model selection criteria to choose lag lengths.
14. Estimate an autoregressive model and choose a suitable lag length.
15. Use AR and ARDL models to compute forecasts, standard errors of forecasts and forecast intervals.
16. Explain what is meant by exponential smoothing. Use exponential smoothing to compute a forecast.
17. Compute delay, interim, and total multipliers for both ARDL and finite distributed lag models.

Keywords

AIC criterion
AR(1) error
AR(p) model
ARDL(p,q) model
autocorrelation
autoregressive distributed lags
autoregressive error
autoregressive model
BIC criterion
correlogram
delay multiplier
distributed lag weight
dynamic models
exponential smoothing
finite distributed lag
forecast error
forecast intervals
forecasting
HAC standard errors
impact multiplier
infinite distributed lag
interim multiplier
lag length
lag operator
lagged dependent variable
LM test
multiplier analysis
nonlinear least squares
out-of-sample forecasts
sample autocorrelations
serial correlation
standard error of forecast error
SC criterion
serial correlation
total multiplier
$T \times R^2$ form of *LM* test
within-sample forecasts

9.1 Introduction

When modeling relationships between variables, the nature of the data that have been collected has an important bearing on the appropriate choice of an econometric model. In particular, it is important to distinguish between cross-section data (data on a number of economic units at a particular point in time) and time-series data (data collected over time on one particular economic unit). Examples of both types of data were given in Chapter 1.5. When we say ‘‘economic units’’ we could be referring to individuals, households, firms, geographical regions, countries, or some other entity on which data is collected. Because cross-section observations on a number of economic units at a given time are often generated by way of a random sample, they are typically uncorrelated. The level of income observed in the Smiths' household, for example, does not affect, nor is it affected by, the level of income in the Jones's household. On the other hand, time-series observations on a given economic unit, observed over a number of time periods, are likely to be correlated. The level of income observed in the Smiths' household in one year is likely to be related to the level of income in the Smiths' household in the year before. Thus, one feature that distinguishes time-series data from cross-section data is the likely correlation between different observations. Our challenges for this chapter include testing for and modeling such correlation.

A second distinguishing feature of time-series data is its natural ordering according to time. With cross-section data there is no particular ordering of the observations that is better or more natural than another. One could shuffle the observations and then proceed with estimation without losing any information. If one shuffles time-series observations, there is a danger of confounding what is their most important distinguishing feature: the possible existence of dynamic relationships between variables. A dynamic relationship is one in which the change in a variable now has an impact on that same variable, or other variables, in one or more future time periods. For example, it is common for a change in the level of an explanatory variable to have behavioral implications for other variables beyond the time period in which it occurred. The consequences of economic decisions that result in changes in economic variables can last a long time. When the income tax rate is increased, consumers have less disposable income, reducing their expenditures on goods and services, which reduces profits of suppliers, which reduces the demand for productive inputs, which reduces the profits of the input suppliers, and so on. The effect of the tax increase ripples through the economy. These effects do not occur instantaneously but are spread, or **distributed**, over future time periods. As shown in Figure 9.1, economic actions or decisions taken at one point in time, t, have effects on the economy at time t, but also at times $t+1$, $t+2$, and so on.

9.1.1 Dynamic Nature of Relationships

Given that the effects of changes in variables are not always instantaneous, we need to ask how to model the dynamic nature of relationships. We begin by recognizing three different ways of doing so.

1. One way is to specify that a dependent variable y is a function of current and past values of an explanatory variable x. That is,

$$y_t = f(x_t, x_{t-1}, x_{t-2}, \cdots) \tag{9.1}$$

   We can think of (y_t, x_t) as denoting the values for y and x in the current period; x_{t-1} means the value of x in the previous period; x_{t-2} is the value of x two periods ago, and so on. For the moment $f(\cdot)$ is used to denote any general function. Later we replace $f(\cdot)$ by a linear function, like those used so far in the book. Equations such as (9.1) say, for example, that the current rate of inflation y_t depends not just on the current interest rate x_t, but also on the rates in previous time periods x_{t-1}, x_{t-2}, ... Turning this interpretation around as in Figure 9.1, it means that a change in the interest rate now will have an impact on inflation now and in future periods; it takes time for the effect of an interest rate change to fully work its way through the economy. Because of the existence of these lagged effects, (9.1) is called a **distributed lag model**.

2. A second way of capturing the dynamic characteristics of time-series data is to specify a model with a **lagged dependent variable** as one of the explanatory variables. For example,

$$y_t = f(y_{t-1}, x_t) \tag{9.2}$$

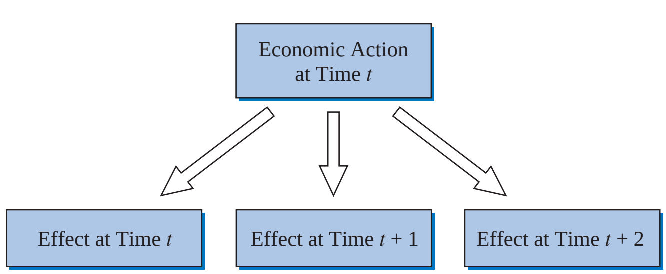

FIGURE 9.1 The distributed lag effect.

where again $f(\cdot)$ is a general function that we later replace with a linear function. In this case we are saying that the inflation rate in one period y_t will depend (among other things) on what it was in the previous period, y_{t-1}. Assuming a positive relationship, periods of high inflation will tend to follow periods of high inflation and periods of low inflation will tend to follow periods of low inflation. Or, in other words, inflation is positively correlated with its value lagged one period. A model of this nature is one way of modeling correlation between current and past values of a dependent variable. Also, we can combine the features of (9.1) and (9.2) so that we have a dynamic model with lagged values of both the dependent and explanatory variables, such as

$$y_t = f(y_{t-1}, x_t, x_{t-1}, x_{t-2}) \tag{9.3}$$

Such models are called **autoregressive distributed lag (ARDL)** models, with "autoregressive" meaning a regression of y_t on its own lag or lags.

3. A third way of modeling the continuing impact of change over several periods is via the error term. For example, using general functions $f(\cdot)$ and $g(\cdot)$, both of which are replaced later with linear functions, we can write

$$y_t = f(x_t) + e_t \qquad e_t = g(e_{t-1}) \tag{9.4}$$

where the function $e_t = g(e_{t-1})$ is used to denote the dependence of the error on its value in the previous period. In this case e_t is correlated with e_{t-1}; we say the errors are **serially correlated** or **autocorrelated**. Because (9.3) implies $e_{t+1} = g(e_t)$, the dynamic nature of this relationship is such that the impact of any unpredictable shock that feeds into the error term will be felt not just in period t, but also in future periods. The current error e_t affects not just the current value of the dependent variable y_t, but also its future values $y_{t+1}, y_{t+2}, \ldots$. As an example, suppose that a terrorist act creates fear of an oil shortage, driving up the price of oil. The terrorist act is an unpredictable shock that forms part of the error term e_t. It is likely to affect the price of oil in the future as well as during the current period.

In this chapter we consider these three ways in which dynamics can enter a regression relationship—lagged values of the explanatory variable, lagged values of the dependent variable, and lagged values of the error term. What we discover is that these three ways are not as distinct as one might at first think. Including a lagged dependent variable y_{t-1} can capture similar effects to those obtained by including a lagged error e_{t-1}, or a long history of past values of an explanatory variable, $x_{t-1}, x_{t-2}, \ldots$. Thus, we not only consider the three kinds of dynamic relationships, we explore the relationships between them.

Related to the idea of modeling dynamic relationships between time series variables is the important concept of forecasting. We are not only interested in tracing the impact of a change in an explanatory variable or an error shock through time. Forecasting future values of economic time series, such as inflation, unemployment, and exchange rates, is something that attracts the attention of business, governments, and the general public. Describing how dynamic models can be used for forecasting is another objective of this chapter.

9.1.2 Least Squares Assumptions

An important consequence of using time series data to estimate dynamic relationships is the possible violation of one of our least squares assumptions. Assumption MR4 specified in Chapter 5 states that different observations on y and on e are uncorrelated. That is,

$$\text{cov}(y_i, y_j) = \text{cov}(e_i, e_j) = 0 \quad \text{for } i \neq j$$

In this chapter, to emphasize that we are using time-series observations, we drop the i and j subscripts and use t and s instead, with t and s referring to two different time periods such as days, months, quarters, or years. Thus, the above assumption becomes

$$\text{cov}(y_t, y_s) = \text{cov}(e_t, e_s) = 0 \text{ for } t \neq s$$

The dynamic models in (9.2), (9.3) and (9.4) imply correlation between y_t and y_{t-1} or e_t and e_{t-1} or both, so they clearly violate assumption MR4. As mentioned below (9.4), when a variable is correlated with its past values, we say that it is autocorrelated or serially correlated. How to test for serial correlation, and its implications for estimation, are also covered in this chapter.

9.1.2a Stationarity

An assumption that we maintain throughout the chapter is that the variables in our equations are **stationary**. This assumption will take on more meaning in Chapter 12 when it is relaxed. For the moment we note that a stationary variable is one that is not explosive, nor trending, and nor wandering aimlessly without returning to its mean. These features can be illustrated with some graphs. Figures 9.2(a), 9.2(b) and 9.2(c) contain graphs of the observations on three different variables, plotted against time. Plots of this kind are routinely considered when examining time-series variables. The variable Y that appears in Figure 9.2(a) is considered stationary because it tends to fluctuate around a constant mean without wandering or trending. On the other hand, X and Z in Figures 9.2(b) and 9.2(c) possess characteristics of nonstationary variables. In Figure 9.2(b) X tends to wander, or is "slow turning," while Z in Figure 9.2(c) is trending. These concepts will be defined more precisely in Chapter 12. For now the important thing to remember is that this chapter is concerned with modeling and estimating dynamic relationships between stationary variables whose time series have similar characteristics to those of Y. That is, they neither wander nor trend.

9.1.3 Alternative Paths Through the Chapter

This chapter covers a great deal of material. Instructors teaching a one-semester course may not wish to cover all of it, and different instructors are likely to have different preferences for the sections they wish to cover. Figures 9.3(a) and 9.3(b) provide a guide to alternative ways of covering a limited amount of the material. Figure 9.3(a) is designed for instructors who wish to start with finite distributed lags. This starting point has the advantage of beginning with a model that is closest to those studied so far in Chapters 2 to 8. From there we recommend covering serial correlation—relevant definitions, concepts, and testing. At this point some instructors might like to proceed with the AR(1) error model; others might prefer to jump straight to ARDL models. The second path in Figure 9.3(b) is designed for instructors who wish to start the chapter with serial correlation. After covering definitions, concepts, and testing, they can proceed to the AR(1) error model or straight to ARDL models. Finite distributed lag models can be covered as a special case of ARDL models or omitted.

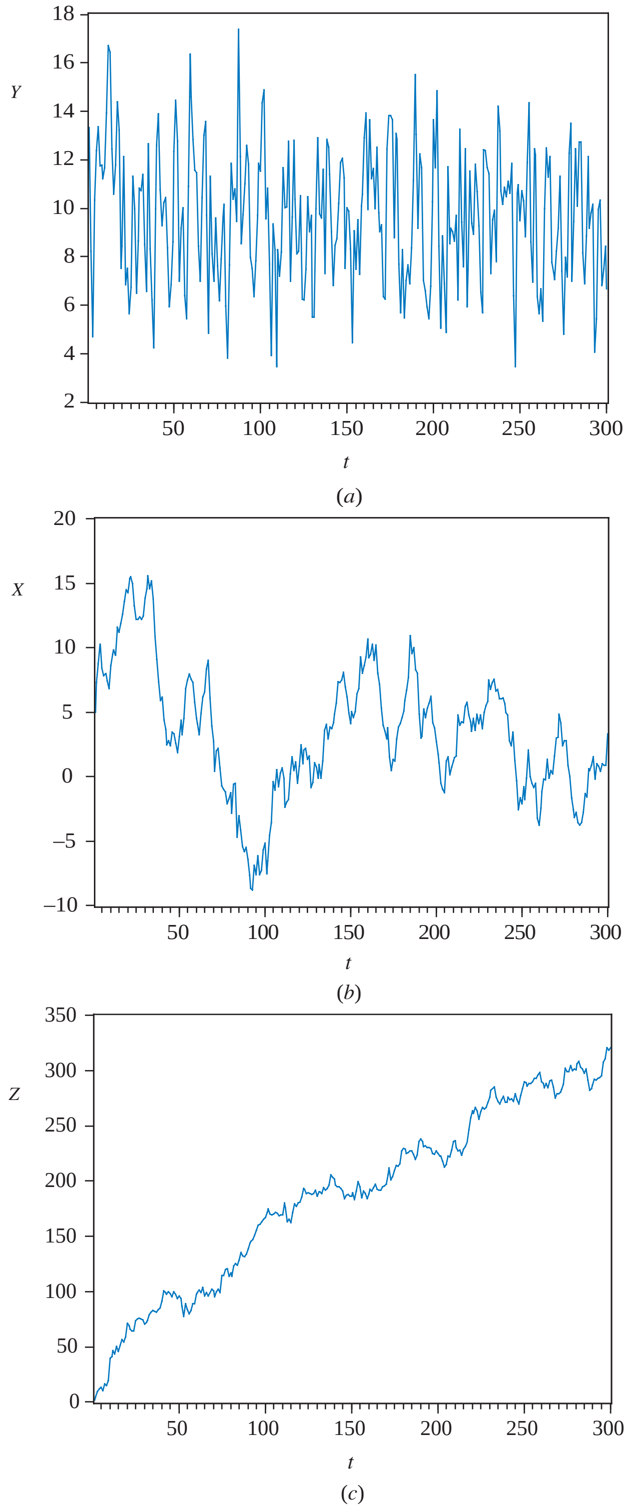

FIGURE 9.2 (a) Time series of a stationary variable; (b) time series of a nonstationary variable that is "slow-turning" or "wandering"; (c) time series of a nonstationary variable that "trends."

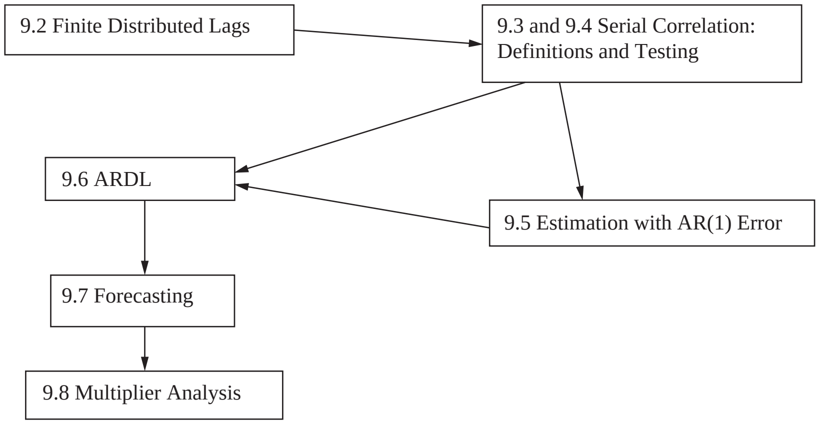

FIGURE 9.3 (a) Alternative paths through the chapter starting with finite distributed lags.

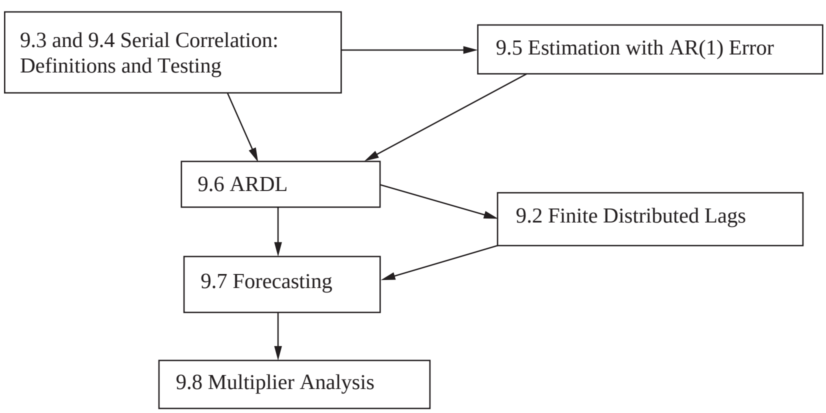

FIGURE 9.3 (b) Alternative paths through the chapter starting with serial correlation.

9.2 Finite Distributed Lags

The first dynamic relationship that we consider is that given in (9.1), $y_t = f(x_t, x_{t-1}, x_{t-2}, \cdots)$, with the additional assumptions that the relationship is linear, and, after q time periods, changes in x no longer have an impact on y. Under these conditions we have the multiple regression model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \tag{9.5}$$

The model in (9.5) can be treated in the same way as the multiple regression model studied in Chapters 5 and 6. Instead of having a number of explanatory variables, we have a number of different lags of the same explanatory variable. However, for the purpose of estimation, these different lags can be treated in the same way as different explanatory variables. It is convenient to change subscript notation on the coefficients: β_s is used to denote the coefficient of x_{t-s} and α is introduced to denote the intercept. Other explanatory variables can be added if relevant, in which case other symbols are needed to denote their coefficients.

Models such as (9.5) have two special uses. The first is **forecasting** future values of y. To introduce notation for future values, suppose our sample period is for $t = 1, 2, \ldots, T$. We use t for the index (rather than i) and T for the sample size (rather than N) to emphasize the time series nature of the data. Given that the last observation in our sample is at $t = T$, the first postsample observation that we want to forecast is at $t = T + 1$. The equation for this observation is given by

$$y_{T+1} = \alpha + \beta_0 x_{T+1} + \beta_1 x_T + \beta_2 x_{T-1} + \cdots + \beta_q x_{T-q+1} + e_{T+1} \tag{9.6}$$

The forecasting problem is how to use the time series of x-values, $x_{T+1}, x_T, x_{T-1}, \ldots, x_{T-q+1}$ to forecast the value y_{T+1}, with special attention needed to obtain a value for x_{T+1}. We consider this problem in Section 9.7, within the context of a more general model.

The second special use of models like (9.5) is for **policy analysis**. Examples of policy analysis where the distributed-lag effect is important are the effects of changes in government expenditure or taxation on unemployment and inflation (fiscal policy), the effects of changes in the interest rate on unemployment and inflation (monetary policy), and the effect of advertising on sales of a firm's products. The timing of the effect of a change in the interest rate or a change in taxation on unemployment, inflation, and the general health of the economy can be critical. Suppose the government (or a firm or business) controls the values of x, and would like to set x to achieve a given value, or a given sequence of values, for y. The coefficient β_s gives the change in $E(y_t)$ when x_{t-s} changes by one unit, but x is held constant in other periods. Alternatively, if we look forward instead of backward, β_s gives the change in $E(y_{t+s})$ when x_t changes by one unit, but x in other periods is held constant. In terms of derivatives

$$\frac{\partial E(y_t)}{\partial x_{t-s}} = \frac{\partial E(y_{t+s})}{\partial x_t} = \beta_s \tag{9.7}$$

To further appreciate this interpretation, suppose that x and y have been constant for at least the last q periods and that x_t is increased by one unit, then returned to its original level. Then, using (9.5) but ignoring the error term, the immediate effect will be an increase in y_t by β_0 units. One period later, y_{t+1} will increase by β_1 units, then y_{t+2} will increase by β_2 units and so on, up to period $t + q$, when y_{t+q} will increase by β_q units. In period $t + q + 1$ the value of y will return to its original level. The effect of a one-unit change in x_t is **distributed** over the current and next q periods, from which we get the term "distributed lag model." It is called a **finite distributed lag model of order *q*** because it is assumed that after a finite number of periods q, changes in x no longer have an impact on y. The coefficient β_s is called a **distributed-lag weight** or an ***s*-period delay multiplier**. The coefficient β_0 ($s = 0$) is called the **impact multiplier**.

It is also relevant to ask what happens if x_t is increased by one unit and then maintained at its new level in subsequent periods $(t + 1), (t + 2), \ldots$. In this case, the immediate impact will again be β_0; the total effect in period $t + 1$ will be $\beta_0 + \beta_1$, in period $t + 2$ it will be $\beta_0 + \beta_1 + \beta_2$, and so on. We add together the effects from the changes in all preceding periods. These quantities are called **interim multipliers**. For example, the two-period interim multiplier is $(\beta_0 + \beta_1 + \beta_2)$. The **total multiplier** is the final effect on y of the sustained increase after q or more periods have elapsed; it is given by $\sum_{s=0}^{q} \beta_s$.

9.2.1 Assumptions

When the simple regression model was first introduced in Chapter 2, it was written in terms of the mean of y conditional on x. Specifically, $E(y|x) = \beta_1 + \beta_2 x$, which led to the error term assumption $E(e|x) = 0$. Then, so that we could avoid the need to condition on x, and hence ease the notational burden, we made the simplifying assumption that the x's are not random. We maintained this assumption through Chapters 2–8, recognizing that although it is unrealistic for most data sets, relaxing it in a limited but realistic way would have had little impact on our results and on our choice of estimators and test statistics. Further consequences of relaxing it are explored in Chapter 10. However, because the time-series variables used in the examples in this chapter are random, it is useful to mention alternative assumptions under which we can consider the properties of least squares and other estimators.

In distributed lag models both y and x are typically random. The variables used in the example that follows are unemployment and output growth. They are both random. They are observed at the same time; we do not know their values prior to "sampling." We do not "set" output growth and then observe the resulting level of unemployment. To accommodate this randomness we assume that the x's are random and that e_t is independent of all x's in the sample—past, current, and future. This assumption, in conjunction with the other multiple regression assumptions, is sufficient for the least squares estimator to be unbiased and to be best linear unbiased conditional on the x's in the sample.[1] With the added assumption of normally distributed error terms, our usual t and F tests have finite sample justification. Accordingly, the multiple regression assumptions given in Chapter 5 can be modified for the distributed lag model as follows:

ASSUMPTIONS OF THE DISTRIBUTED LAG MODEL

TSMR1. $y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \qquad t = q+1, \ldots, T$

TSMR2. y and x are stationary random variables, and e_t is independent of current, past and future values of x.

TSMR3. $E(e_t) = 0$

TSMR4. $\text{var}(e_t) = \sigma^2$

TSMR5. $\text{cov}(e_t, e_s) = 0 \quad t \neq s$

TSMR6. $e_t \sim N(0, \sigma^2)$

9.2.2 An Example: Okun's Law

To illustrate and expand on the various distributed lag concepts, we introduce an economic model known as Okun's Law. In this model the change in the unemployment rate from one period to the next depends on the rate of growth of output in the economy:[2]

$$U_t - U_{t-1} = -\gamma(G_t - G_N) \tag{9.8}$$

[1] The complete independence of e and all x's is stronger than needed to establish good large sample properties. See Section 9.5 and Chapter 10.

[2] See O. Blanchard (2009), *Macroeconomics*, 5th edition, Upper Saddle River, NJ, Pearson Prentice Hall, p. 184.

where U_t is the unemployment rate in period t, G_t is the growth rate of output in period t, and G_N is the "normal" growth rate, which we assume is constant over time. The parameter γ is positive, implying that when the growth of output is above the normal rate, unemployment falls; a growth rate below the normal rate leads to an increase in unemployment. The normal growth rate G_N is the rate of output growth needed to maintain a constant unemployment rate. It is equal to the sum of labor force growth and labor productivity growth. We expect $0 < \gamma < 1$, reflecting that output growth leads to less than one-to-one adjustments in unemployment.[3]

To write (9.8) in the more familiar notation of the multiple regression model, we denote the change in unemployment by $DU_t = \Delta U_t = U_t - U_{t-1}$,[4] we set $\beta_0 = -\gamma$, and $\alpha = \gamma G_N$. Including an error term then yields

$$DU_t = \alpha + \beta_0 G_t + e_t \tag{9.9}$$

Recognizing that changes in output are likely to have a distributed-lag effect on unemployment—not all of the effect will take place instantaneously—we expand (9.9) to include lags of G_t

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \cdots + \beta_q G_{t-q} + e_t \tag{9.10}$$

To estimate this relationship we use quarterly U.S. data on unemployment and the percentage change in gross domestic product (GDP) from quarter 2, 1985, to quarter 3, 2009. Output growth is defined as

$$G_t = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}} \times 100 \tag{9.11}$$

These data are stored in the file *okun.dat*. The time series for DU and G are graphed in Figures 9.4(a) and 9.4(b). The effects of the global financial crisis are clearly evident towards the end of the sample. At this time we note that the series appear to be stationary; tools for more rigorous assessment of stationarity are deferred until Chapter 12.

To fully appreciate how the lagged variables are defined and how their observations enter the estimation procedure, consider the spreadsheet in Table 9.1. This table contains the observations on U_t, its lag U_{t-1}, and its difference DU_t, as well as G_t and its lags up to G_{t-3}. Notice that for $t = 2$, $U_t = U_2 = 7.2$, $U_{t-1} = U_1 = 7.3$, and $DU_t = U_2 - U_1 = 7.2 - 7.3 = -0.1$. Similarly, for $t = 3$, $U_t = U_3 = 7.0$, $U_{t-1} = U_2 = 7.2$, and $DU_t = U_3 - U_2 = 7.0 - 7.2 = -0.2$. No observations are listed for U_{t-1} and DU_t for $t = 1$, because they would require a value for U_0 (1985Q1) which is not provided in this data set. For G_t, when $t = 2$, $G_{t-1} = G_1 = 1.4$. When $t = 3$, $G_{t-1} = G_2 = 2.0$ and $G_{t-2} = G_1 = 1.4$. When $t = 4$, $G_{t-1} = G_3 = 1.4$, $G_{t-2} = G_2 = 2.0$, and $G_{t-3} = G_1 = 1.4$. Because an observation for G_0 is not available, an observation is lost for each lag that is introduced. Using three lags of G ($q = 3$) means that only 95 of the original 98 observations are used for estimation.[5] In the general case with q lags, the observations used are those for $t = q + 1, \; q + 2, \ldots, T$.

[3] For more details see Blanchard (2009), *ibid*, Chapter 9.

[4] Using DU_t, instead of U_t, has two advantages. The first is that Okun's Law is stated in terms of the change in unemployment. The second is that DU_t, is stationary, but U_t, is not.

[5] Since G is defined as the percentage change in GDP, one might question whether an extra observation on G should be lost. However, the growth rate for 1985Q2 was obtained directly from Federal Reserve economic data.

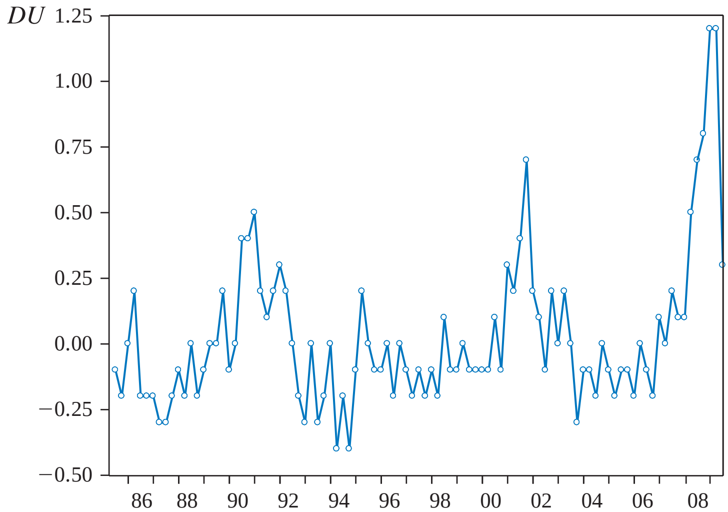

FIGURE 9.4 (a) Time series for the change in the U.S. unemployment rate: 1985Q3 to 2009Q3.

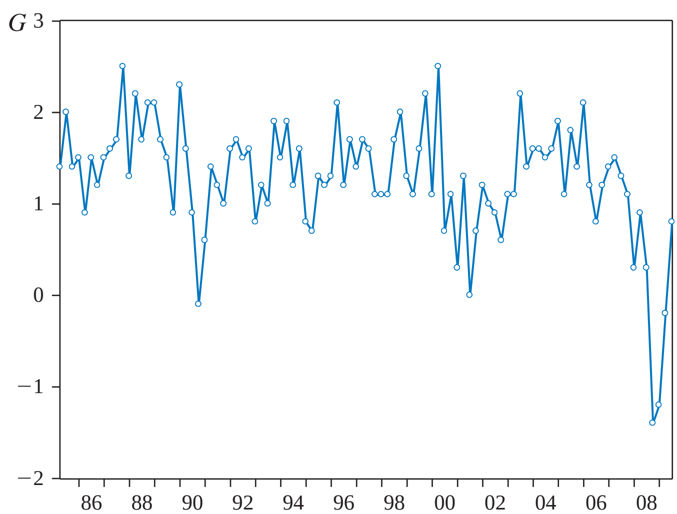

FIGURE 9.4 (b) Time series for U.S. GDP growth: 1985Q2 to 2009Q3.

Table 9.1 **Spreadsheet of Observations for Distributed Lag Model**

t	Quarter	U_t	U_{t-1}	DU_t	G_t	G_{t-1}	G_{t-2}	G_{t-3}
1	1985Q2	7.3	•	•	1.4	•	•	•
2	1985Q3	7.2	7.3	−0.1	2.0	1.4	•	•
3	1985Q4	7.0	7.2	−0.2	1.4	2.0	1.4	•
4	1986Q1	7.0	7.0	0.0	1.5	1.4	2.0	1.4
5	1986Q2	7.2	7.0	0.2	0.9	1.5	1.4	2.0
96	2009Q1	8.1	6.9	1.2	−1.2	−1.4	0.3	0.9
97	2009Q2	9.3	8.1	1.2	−0.2	−1.2	−1.4	0.3
98	2009Q3	9.6	9.3	0.3	0.8	−0.2	−1.2	−1.4

Least squares estimates of the coefficients and related statistics for (9.10) are reported in Table 9.2 for lag lengths $q = 2$ and $q = 3$. All coefficients of G and its lags have the expected negative sign and are significantly different from zero at a 5% significance level, with the exception of that for G_{t-3} when $q = 3$. A variety of measures are available for choosing q. In this case we drop G_{t-3} and settle on a model of order 2 because b_3 is insignificant and has the wrong sign, and b_0, b_1, and b_2 all have the expected negative signs and are significantly different from zero. The information criteria AIC and SC discussed in Chapter 6 are another set of measures that can be used for assessing lag length.

What do the estimates for lag length 2 tell us? A 1% increase in the growth rate leads to a fall in the unemployment rate of 0.20% in the current quarter, a fall of 0.16% in the next quarter, and a fall of 0.07% two quarters from now, holding other factors fixed. These changes represent the values of the impact multiplier and the one-quarter and two-quarter delay multipliers. The interim multipliers, that give the effect of a sustained increase in the growth rate of 1%, are -0.367 for one quarter and -0.437 for two quarters. Since we have a lag length of two, -0.437 is also the total multiplier. Knowledge of these values is important for a government who wishes to keep unemployment below a certain level by influencing the growth rate. If we view γ in (9.8) as the total effect of a change in output growth, then its estimate is $\hat{\gamma} = -\sum_{s=0}^{2} b_s = 0.437$. An estimate of the normal growth rate that is needed to maintain a constant unemployment rate is $\hat{G}_N = \hat{\alpha}/\hat{\gamma} = 0.5836/0.437 = 1.3\%$ per quarter.

A possibly puzzling result in Table 9.2 is that the estimated model with G_{t-3} has a slightly lower R^2 than that without G_{t-3}. Since adding a variable lowers the sum of squared errors and increases the R^2, this outcome is counterintuitive. It can occur in this case because the number of observations is different in each case. If we are using all of the data available, the number of observations changes as the number of lags changes unless specific provision is made to do otherwise.

Table 9.2 Estimates for Okun's Law Finite Distributed Lag Model

Lag Length $q = 3$				
Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5810	0.0539	10.781	0.0000
G_t	-0.2021	0.0330	6.120	0.0000
G_{t-1}	-0.1645	0.0358	-4.549	0.0000
G_{t-2}	-0.0716	0.0353	-2.027	0.0456
G_{t-3}	0.0033	0.0363	0.091	0.9276
Observations = 95	$R^2 = 0.652$		$\hat{\sigma} = 0.1743$	

Lag Length $q = 2$				
Variable	Coefficient	Std. Error	t-value	p-value
Constant	0.5836	0.0472	12.360	0.0000
G_t	-0.2020	0.0324	-6.238	0.0000
G_{t-1}	-0.1653	0.0335	-4.930	0.0000
G_{t-2}	-0.0700	0.0331	-2.115	0.0371
Observations = 96	$R^2 = 0.654$		$\hat{\sigma} = 0.1726$	

9.3 Serial Correlation

In the distributed lag model in the previous section we examined one feature of time-series data: how a dependent variable can be related to current and past values of an explanatory variable. The effect of a change in the value of an explanatory variable is distributed over a number of future periods. We noted that, if the specified assumptions hold, and, in particular, the equation errors are uncorrelated with each other and with x, the traditional least squares estimator and associated testing procedures can be used.

We turn now to another question: When is assumption TSMR5, $\text{cov}(e_t, e_s) = 0$ for $t \neq s$ likely to be violated, and how do we assess its validity? As mentioned in the introduction to this chapter, different observations in a cross-section data set, collected by way of a random sample, are typically uncorrelated. With time-series data, however, successive observations are likely to be correlated. If unemployment is high in this quarter, it is more likely to be high than low next quarter. Changes in variables such as unemployment, output growth, inflation and interest rates are usually more gradual than abrupt; their values in one period will depend on what happened in the previous period. This dependence means that output growth now, for example, will be correlated with output growth in the previous period. When a variable exhibits correlation over time, we say it is **autocorrelated** or **serially correlated**; we will use these two terms interchangeably. Both observable time-series variables such as DU and G, and the unobservable error e, can be autocorrelated. Autocorrelation in the error can arise from an autocorrelated omitted variable, or it can arise if a dependent variable y is autocorrelated and this autocorrelation is not adequately explained by the x's and their lags that are included in the equation.

To illustrate the concept of autocorrelation or serial correlation, we begin by considering the observations on output growth G that were used in the distributed lag model of the previous section. We describe methodology for measuring autocorrelation and for testing whether it is significantly different from zero. Then, later in this section, we apply the methodology to the error term in a regression equation. It is useful to assess the autocorrelation properties of both observable variables and the error term. For the observable variables, the properties are useful for the construction of autoregressive models that are considered later in this chapter. For the error term it is useful to check whether one of the least squares assumptions has been violated.

9.3.1 Serial Correlation in Output Growth

To appreciate the nature of autocorrelation, consider the time-series graph of G in Figure 9.4(b). In a few instances G changes dramatically from one quarter to the next, but on average, high values of G_{t-1} are followed by high values of G_t, and low values of G_{t-1} are followed by low values of G_t, suggesting a positive correlation between observations that are one period apart. We can further illustrate this correlation by examining the scatter diagram in Figure 9.5 where pairs of observations (G_{t-1}, G_t) are plotted using the data from Table 9.1.[6] If G_t and G_{t-1} are uncorrelated, the observations would be scattered randomly throughout all four quadrants. The predominance of points in the NE and SW quadrants suggests G_t and G_{t-1} are positively correlated.

[6] This diagrammatic tool was introduced in Figure P.4 in the Probability Primer to explain the meaning of covariance and correlation.

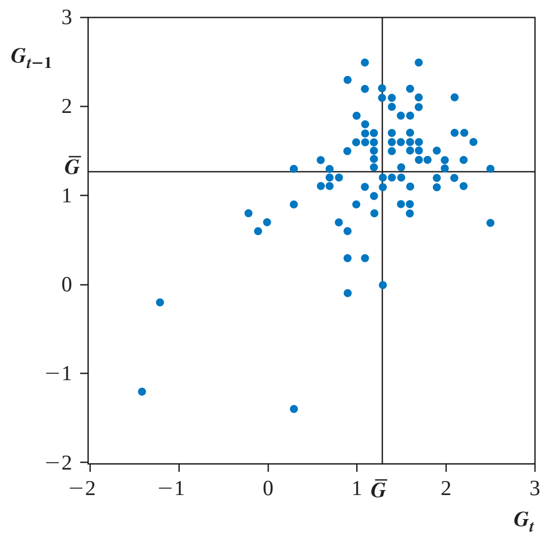

FIGURE 9.5 Scatter diagram for G_t and G_{t-1}.

9.3.1a Computing Autocorrelations

The correlations between a variable and its lags are called autocorrelations. How do we measure this kind of correlation? Recall from Chapter 4.2 that the population correlation between two variables x and y is given by

$$\rho_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}}$$

Turning this formula into one that measures the correlation between G_t and G_{t-1}, we have

$$\rho_1 = \frac{\operatorname{cov}(G_t, G_{t-1})}{\sqrt{\operatorname{var}(G_t)\operatorname{var}(G_{t-1})}} = \frac{\operatorname{cov}(G_t, G_{t-1})}{\operatorname{var}(G_t)} \tag{9.12}$$

The notation ρ_1 is used to denote the population correlation between observations that are one period apart in time, known also as the **population autocorrelation of order one**. The second equality in (9.12) holds because $\operatorname{var}(G_t) = \operatorname{var}(G_{t-1})$, a property of time series that are stationary.

The **first-order sample autocorrelation** for G is obtained from (9.12) by replacing $\operatorname{cov}(G_t, G_{t-1})$ and $\operatorname{var}(G_t)$ by their estimates

$$\widehat{\operatorname{cov}(G_t, G_{t-1})} = \frac{1}{T-1}\sum_{t=2}^{T}\left(G_t - \overline{G}\right)\left(G_{t-1} - \overline{G}\right), \qquad \widehat{\operatorname{var}(G_t)} = \frac{1}{T-1}\sum_{t=1}^{T}\left(G_t - \overline{G}\right)^2$$

where $\overline{G}$ is the sample mean $\overline{G} = T^{-1}\sum_{t=1}^{T} G_t$. The index of summation in the formula for $\widehat{\operatorname{cov}(G_t, G_{t-1})}$ starts at $t = 2$ because we do not observe G_0. Making the substitutions, and using r_1 to denote the sample autocorrelation at lag one, yields

$$r_1 = \frac{\sum_{t=2}^{T}\left(G_t - \overline{G}\right)\left(G_{t-1} - \overline{G}\right)}{\sum_{t=1}^{T}\left(G_t - \overline{G}\right)^2} \tag{9.13}$$

More generally, **the k-th order sample autocorrelation** for a series y that gives the correlation between observations that are k periods apart (the correlation between y_t and y_{t-k}) is given by

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2} \tag{9.14}$$

This formula is commonly used in the literature and in software and is the one we use to compute autocorrelations in this text, but it is worth mentioning variations of it that are sometimes used. Because $(T - k)$ observations are used to compute the numerator and T observations are used to compute the denominator, an alternative that leads to larger estimates in finite samples is

$$r'_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2} \tag{9.15}$$

Another modification of (9.14) that has a similar effect is to use only $(T - k)$ observations in the denominator, so that it becomes $\sum_{t=k+1}^{T}(y_t - \bar{y})^2$.

Applying (9.14) to the series G yields, for the first four autocorrelations,

$$r_1 = 0.494 \quad r_2 = 0.411 \quad r_3 = 0.154 \quad r_4 = 0.200 \tag{9.16}$$

The autocorrelations at lags one and two are moderately high; those at lags three and four are much smaller—less than half the magnitude of the earlier ones. How do we test whether an autocorrelation is significantly different from zero? Let the kth order population autocorrelation be denoted by ρ_k. Then, when the null hypothesis $H_0 : \rho_k = 0$ is true, it turns out that r_k has an approximate normal distribution with mean zero and variance $1/T$. Thus, a suitable test statistic is

$$Z = \frac{r_k - 0}{\sqrt{1/T}} = \sqrt{T}r_k \sim N(0, 1) \tag{9.17}$$

The product of the square root of the sample size and the sample autocorrelation r_k has an approximate standard normal distribution. At a 5% significance level, we reject $H_0 : \rho_k = 0$ when $\sqrt{T}r_k \geq 1.96$ or $\sqrt{T}r_k \leq -1.96$.

For the series G, $T = 98$, and the values of the test statistic Z for the first four lags are

$$\begin{aligned} Z_1 &= \sqrt{98} \times 0.494 = 4.89, \qquad & Z_2 &= \sqrt{98} \times 0.414 = 4.10 \\ Z_3 &= \sqrt{98} \times 0.154 = 1.52, \qquad & Z_4 &= \sqrt{98} \times 0.200 = 1.98 \end{aligned}$$

Thus, we reject the hypotheses $H_0 : \rho_1 = 0$ and $H_0 : \rho_2 = 0$, we have insufficient evidence to reject $H_0 : \rho_3 = 0$, and r_4 is on the borderline of being significant. We conclude that G, the quarterly growth rate in U.S. GDP, exhibits significant serial correlation at lags one and two.

9.3.1b The Correlogram

A useful device for assessing the significance of autocorrelations is a diagrammatic representation of the correlogram. The **correlogram,** also called the **sample autocorrelation function**, is the sequence of autocorrelations $r_1,\ r_2,\ r_3, \ldots$. It shows the correlation between observations that are one period apart, two periods apart, three periods apart, and so on. We indicated that an autocorrelation r_k will be significantly different from zero at a 5% significance level if $\sqrt{T}r_k \geq 1.96$ or if $\sqrt{T}r_k \leq -1.96$. Alternatively, we can say that r_k will be significantly different from zero if $r_k \geq 1.96/\sqrt{T}$ or $r_k \leq -1.96/\sqrt{T}$. By drawing the values $\pm 1.96/\sqrt{T}$ as bounds on a graph that illustrates the magnitude of each of the r_k, we can see at a glance which correlations are significant.

A graph of the correlogram for G for the first 12 lags appears in Figure 9.6. The heights of the bars represent the correlations and the horizontal lines drawn at $\pm 2/\sqrt{98} = \pm 0.202$ are the significance bounds. We have used 2 rather than 1.96 as a convenient approximation. We can see at a glance that r_1 and r_2 are significantly different from zero, that r_4 and r_{12} are bordering on significance, and the remainder of the autocorrelations are not significantly different from zero.

Your software may not produce a correlogram that is exactly the same as Figure 9.6. It might have the correlations on the x-axis and the lags on the y-axis. It could use spikes instead of bars to denote the correlations, it might provide a host of additional information, and the width of its significance bounds might vary with different lags. So be prepared! Learn to isolate and focus on the information corresponding to that in Figure 9.5 and do not be disturbed if the output is slightly different. If the significance bounds vary, it is because they use a refinement of the large sample approximation $\sqrt{T}r_k \sim N(0, 1)$.

Before turning to the question of autocorrelated errors in a regression equation, we note a few facts related to the stationarity of a series. The formula used for computing autocorrelations assumes that the mean and variance of the series are constant over time, and that an autocorrelation depends on the time between observations, not on the actual time period. These are characteristics of a stationary time series—characteristics that are more precise than our earlier vague description of a stationary time series as one which neither trends nor wanders. These issues are explored in detail in Chapter 12.

9.3.2 Serially Correlated Errors

The correlogram can also be used to check whether the multiple regression assumption $\text{cov}(e_t, e_s) = 0$ for $t \neq s$ is violated. To illustrate how to do so, we introduce another

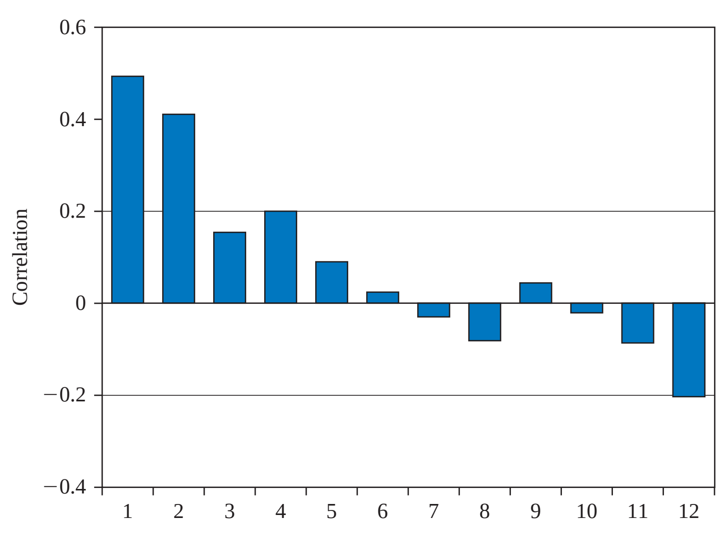

FIGURE 9.6 Correlogram for G.

example: the Phillips curve. A further test for serially correlated errors is considered in Section 9.4. In Section 9.5 we investigate the consequences of serial correlation for least squares estimates, and examine alternative ways of overcoming the problem. Autoregressive distributed lag models, which provide a very general way of allowing for serially correlated errors and at the same time accommodate the dynamic features of lagged y's and lagged x's, are considered in Section 9.6, where we re-examine both the Phillips curve and the model for Okun's Law.

9.3.2a A Phillips Curve

The Phillips curve has a long history in macroeconomics as a tool for describing the relationship between inflation and unemployment.[7] Our starting point is the model

$$INF_t = INF_t^E - \gamma(U_t - U_{t-1}) \tag{9.18}$$

where INF_t is the inflation rate in period t, INF_t^E denotes inflationary expectations for period t, $DU_t = U_t - U_{t-1}$ denotes the change in the unemployment rate from period $t-1$ to period t, and γ is an unknown positive parameter. It is hypothesized that falling levels of unemployment $(U_t - U_{t-1} < 0)$ reflect excess demand for labor that drives up wages, which in turn drives up prices. Conversely, rising levels of unemployment $(U_t - U_{t-1} > 0)$ reflect an excess supply of labor that moderates wage and price increases. The expected inflation rate is included because workers will negotiate wage increases to cover increasing costs from expected inflation, and these wage increases will be transmitted into actual inflation. We initially assume that inflationary expectations are constant over time $(\beta_1 = INF_t^E)$, an assumption that we relax in Section 9.5. With this change, setting $\beta_2 = -\gamma$, and adding an error term, the Phillips curve can be written as the simple regression model

$$INF_t = \beta_1 + \beta_2 DU_t + e_t \tag{9.19}$$

The data used for estimating (9.19) are quarterly Australian data from 1987, Quarter 1 to 2009, Quarter 3. The data are stored in the file *phillips_aus.dat*. One observation is lost in the construction of $DU_t = U_t - U_{t-1}$, making the sample period from 1987Q2 to 2009Q3, a total of 90 observations. Inflation is calculated as the percentage change in the Consumer Price Index, with an adjustment in the third quarter of 2000 when Australia introduced a national sales tax. The adjusted time series is graphed in Figure 9.7(a); the time series for the change in the unemployment rate appears in Figure 9.7(b). While both of these graphs wander a bit, we will proceed under the assumption that they are stationary. More formal tests are set as exercises in Chapter 12.

To examine whether the errors in the Phillips curve in (9.19) are serially correlated, we first compute the least squares residuals

$$\hat{e}_t = INF_t - b_1 - b_2 DU_t \tag{9.20}$$

Because the e_t are unobserved, it is impossible to compute their autocorrelations. We rely instead on the correlogram of the residuals which is an estimate of the correlogram of the unobserved errors and hence provides evidence on whether or not the assumption

[7] For a historical review of the development of different versions, see Gordon, R. J. (2008), "The History of the Phillips Curve: An American Perspective," www.nzae.org.nz/conferences/2008/090708/nr1217302437.pdf. Keynote Address at the Australasian Meetings of the Econometric Society.

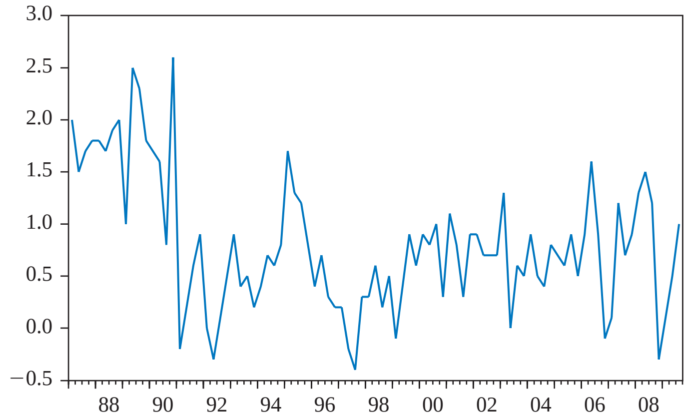

FIGURE 9.7 (a) Time series for Australian price inflation.

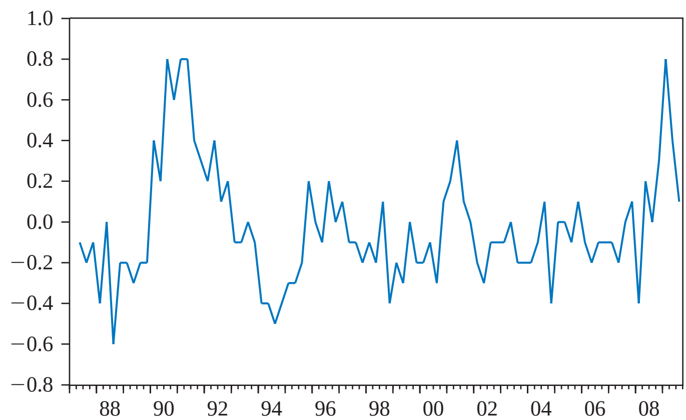

FIGURE 9.7 (b) Time series for the quarterly change in the Australian unemployment rate.

$\text{cov}(e_t, e_s) = 0$ is violated. Replacing y by $\hat{e}$ in (9.14), and recalling that the sample mean of the least squares residuals is zero, the k-th order autocorrelation for the residuals can be written as

$$r_k = \frac{\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^{T} \hat{e}_t^2} \tag{9.21}$$

The least squares estimated equation is

$$\begin{aligned} \widehat{INF} &= 0.7776 - 0.5279DU \\ (\text{se}) \quad & \ (0.0658) \quad (0.2294) \end{aligned} \tag{9.22}$$

These preliminary estimates suggest that an increase in unemployment has the expected negative effect on inflation, and the estimate is significantly different from zero at a 5% significance level.

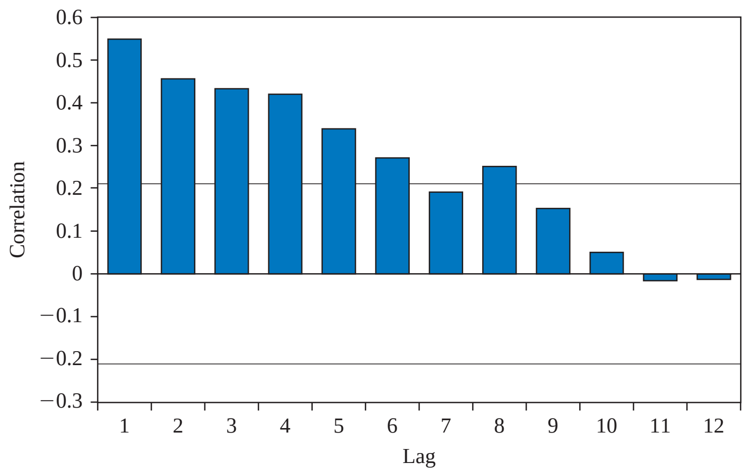

FIGURE 9.8 Correlogram for residuals from least-squares estimated Phillips curve.

Applying (9.21) to the residuals of the least-squares estimated equation in (9.22) yields the correlogram in Figure 9.8. Its significance bounds are $\pm 2/\sqrt{90} = 0.21$. There is strong evidence that the residuals are autocorrelated. The correlations at lags one through six and at lag eight are all significantly different from zero. The values at the first five lags are

$$r_1 = 0.549 \qquad r_2 = 0.456 \qquad r_3 = 0.433 \qquad r_4 = 0.420 \qquad r_5 = 0.339$$

We have found that the errors in the Phillips curve (9.19) are serially correlated. The implications of this correlation and what to do about it are considered in Sections 9.5 and 9.6. Before turning to these solutions, we consider two other tests for serially correlated errors.

9.4 Other Tests For Serially Correlated Errors

9.4.1 A Lagrange Multiplier Test

A second test that we consider for testing for serially correlated errors is derived from a general set of hypothesis testing principles that produce Lagrange[8] multiplier (*LM*) tests. In more advanced courses you will learn the origin of the term Lagrange multiplier. The general principle is described in Appendix C.8.4. An advantage of this test is that it readily generalizes to a **joint** test of correlations at more than one lag.

To introduce the test, suppose in the first instance that we want to test whether errors that are one period apart are correlated. In other words, is $\text{cov}(e_t, e_{t-1})$ equal to zero? Or is r_1 significantly different from zero? If e_t and e_{t-1} are correlated, then one way to model the relationship between them is to write

$$e_t = \rho e_{t-1} + v_t \tag{9.23}$$

[8] Joseph-Louis Lagrange (1736–1813) was an Italian born mathematician. Statistical tests using the so-called Lagrange multiplier principle were introduced into statistics by C. R. Rao in 1948.

where ρ is an unknown parameter and v_t is another random error term. We are saying that e_t depends on e_{t-1}, just like y depends on x in a regression equation where y and x are correlated. Now, if the equation of interest is the simple regression equation $y_t = \beta_1 + \beta_2 x_t + e_t$, then we can substitute (9.23) for e_t, which leads to the equation

$$y_t = \beta_1 + \beta_2 x_t + \rho e_{t-1} + v_t \tag{9.24}$$

Assuming that v_t is independent of e_{t-1}, one way to test whether e_t and e_{t-1} are correlated is to test the null hypothesis $H_0: \rho = 0$. The obvious way to perform this test if e_{t-1} was observable is to regress y_t on x_t and e_{t-1} and to then use a t- or F-test to test the significance of the coefficient of e_{t-1}. However, because e_{t-1} is not observable, we replace it by the lagged least squares residuals $\hat{e}_{t-1}$ and then perform the test in the usual way.

Proceeding in this way seems straightforward, but to complicate matters, time-series econometricians have managed to do it in at least four different ways! Let $t = 1, 2, \ldots, 90$ denote the observations used to estimate the Phillips curve for the period from 1987Q2 to 2009Q3. Then, estimation of (9.24) requires a value for $\hat{e}_0$. Two common ways of overcoming the unavailability of $\hat{e}_0$ are (i) to delete the first observation and hence use a total of 89 observations and (ii) to set $\hat{e}_0 = 0$ and use all 90 observations. The results for the Phillips curve from these two alternatives are

(i)	$t = 6.219$	$F = 38.67$	p-value $= 0.000$
(ii)	$t = 6.202$	$F = 38.47$	p-value $= 0.000$

The results are almost identical. The null hypothesis $H_0: \rho = 0$ is rejected at all conventional significance levels. We conclude that the errors are serially correlated.

As we discovered in Chapter 8, *LM* tests are such that they can frequently be written as the simple expression $T \times R^2$ where T is the number of sample observations and R^2 is the goodness-of-fit statistic from an auxiliary regression. To derive the relevant auxiliary regression for the autocorrelation *LM* test, we begin by writing the test equation from (9.24) as

$$y_t = \beta_1 + \beta_2 x_t + \rho \hat{e}_{t-1} + v_t \tag{9.25}$$

Noting that $y_t = b_1 + b_2 x_t + \hat{e}_t$, we can rewrite (9.25) as

$$b_1 + b_2 x_t + \hat{e}_t = \beta_1 + \beta_2 x_t + \rho \hat{e}_{t-1} + v_t$$

Rearranging this equation yields

$$\begin{aligned} \hat{e}_t &= (\beta_1 - b_1) + (\beta_2 - b_2)x_t + \rho \hat{e}_{t-1} + v_t \\ &= \gamma_1 + \gamma_2 x_t + \rho \hat{e}_{t-1} + v_t \end{aligned} \tag{9.26}$$

where $\gamma_1 = \beta_1 - b_1$ and $\gamma_2 = \beta_2 - b_2$. When testing for autocorrelation by testing the significance of the coefficient of $\hat{e}_{t-1}$, one can estimate (9.25) or (9.26). Both yield the same test result—the same coefficient estimate for $\hat{e}_{t-1}$ and the same t-value. The estimates for the intercept and the coefficient of x_t will be different, however, because in (9.26) we are estimating $(\beta_1 - b_1)$ and $(\beta_2 - b_2)$ instead of β_1 and β_2. The auxiliary regression from which the $T \times R^2$ version of the *LM* test is obtained is (9.26). Because $(\beta_1 - b_1)$ and $(\beta_2 - b_2)$ are centered around zero, if (9.26) is a regression with significant explanatory power, that power will come from $\hat{e}_{t-1}$.

If $H_0: \rho = 0$ is true, then $LM = T \times R^2$ has an approximate $\chi^2_{(1)}$ distribution where T and R^2 are the sample size and goodness-of-fit statistic, respectively, from least squares estimation of (9.26). Once again there are two alternatives depending on whether the first observation is discarded, or $\hat{e}_0$ is set equal to zero. Labeling these two alternatives as (iii) and (iv), respectively, we obtain the following results for the Phillips curve:

(iii) $\quad LM = (T-1) \times R^2 = 89 \times 0.3102 = 27.61$

(iv) $\quad LM = T \times R^2 = 90 \times 0.3066 = 27.59$

These values are much larger than 3.84, which is the 5% critical value from a $\chi^2_{(1)}$-distribution, leading us to reject the null hypothesis of no autocorrelation. Alternatively, we can reject H_0 by examining the p-value for $LM = 27.61$, which is 0.000.

There is no strong theoretical reason for choosing between the four representations of the test. The best one for you to use is that which is automatically computed by your software. We have described all four so that there will not be any mysteries in your computer output.

9.4.1a Testing Correlation at Longer Lags

The correlogram in Figure 9.8 suggested not just correlation between e_t and e_{t-1}, but also between e_t and $(e_{t-2}, e_{t-3}, e_{t-4}, e_{t-5}, e_{t-6})$. The LM test can be used to test for more complicated autocorrelation structures involving higher order lags by including the additional lagged errors in (9.25) or (9.26). An F-test can be used to test the relevance of their inclusion, or, a χ^2-test can be used for the $T \times R^2$ version of the test. The degrees of freedom for the χ^2-test and the numerator degrees of freedom for the F-test are the number of lagged residuals that are included. Slightly different results are obtained depending on whether one discards the initial observations where the lagged values of $\hat{e}_t$ are not available, or sets these values equal to zero. Suppose for the Phillips curve that we add $\hat{e}_{t-2}$, $\hat{e}_{t-3}$ and $\hat{e}_{t-4}$ to (9.26) and use the $T \times R^2$ version of the test. The results we obtain for (iii) omitting the first four observations and (iv) setting $e_0 = e_{-1} = e_{-2} = e_{-3} = 0$ are

(iii) $\quad LM = (T-4) \times R^2 = 86 \times 0.3882 = 33.4$

(iv) $\quad LM = T \times R^2 = 90 \times 0.4075 = 36.7$

Because the 5% critical value from a $\chi^2_{(4)}$- distribution is 9.49, these LM values lead us to conclude that the errors are serially correlated.

9.4.2 The Durbin-Watson Test

The sample correlogram and the Lagrange multiplier test are two large sample tests for serially correlated errors. Their test statistics have their specified distributions in large samples. An alternative test, one that is exact in the sense that its distribution does not rely on a large sample approximation, is the Durbin-Watson test. It was developed in 1950 and for a long time was the standard test for $H_0: \rho = 0$ in the error model $e_t = \rho e_{t-1} + v_t$. It is used less frequently today because its critical values are not available in all software packages, and one has to examine upper and lower critical bounds instead. Also, unlike the LM and correlogram tests, its distribution no longer holds when the equation contains a lagged dependent variable. Details are provided in Appendix 9A at the end of this chapter.

9.5 Estimation With Serially Correlated Errors

In the last two sections we described hypothesis tests for checking whether the least squares assumption $\text{cov}(e_t, e_s) = 0$ is violated. If it is violated, we say the errors are serially correlated. We now ask: What are the implications of serially correlated errors for least squares estimation? And how do we overcome the negative consequences of serially correlated errors? Three estimation procedures are considered:

1. Least squares estimation (Section 9.5.1)
2. An estimation procedure that is relevant when the errors are assumed to follow what is known as a first-order autoregressive model $e_t = \rho e_{t-1} + v_t$ (Section 9.5.2)
3. A general estimation strategy for estimating models with serially correlated errors

The general estimation strategy that will be introduced in Section 9.5.3 and considered in more depth in Section 9.6 is the estimation of an autoregressive distributed lag (ARDL) model which is designed to capture dynamics from all sources—lagged x's, lagged y's, and serially correlated errors.

Before considering each of the above estimation procedures, we need to introduce an extra assumption. We will encounter models with a lagged dependent variable, such as

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

In such cases the time-series assumption TSMR2 introduced in Section 9.2.1 is no longer valid. In the context of the above equation, this assumption says that v_t is not correlated with current, past, and future values of $y_{t-1},\ x_t$ and x_{t-1}. Since y_t is a future value of y_{t-1} and y_t depends directly on v_t, the assumption will be violated. We can, however, replace it with a weaker, more tenable assumption—namely, that v_t is uncorrelated with current and past values of the right-hand-side variables. Under this assumption, the least squares estimator is no longer unbiased, but it does have the desirable large sample property of consistency,[9] and, if the errors are normally distributed, it is best in a large sample sense. Thus, we replace TSMR2 with the following assumption.

> **ASSUMPTION FOR MODELS WITH A LAGGED DEPENDENT VARIABLE**
> TSMR2A: In the multiple regression model $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + v_t$ where some of the x_{tk} may be lagged values of y, v_t is uncorrelated with all x_{tk} and their past values.

This assumption is the one maintained throughout the remainder of this chapter. Note that the v_t are assumed to be uncorrelated random errors with zero mean and constant variance and hence satisfy assumptions TSMR3, TSMR4, and TSMR5 that were previously written in terms of e_t.

[9] The property of consistency is discussed in Appendix 5B.

9.5.1 Least Squares Estimation

First, suppose we proceed with least squares estimation without recognizing the existence of serially correlated errors. What are the consequences? They are essentially the same as ignoring heteroskedasticity should it exist.

1. The least squares estimator is still a linear unbiased estimator, but it is no longer best. If we are able to correctly model the autocorrelated errors, then there exists an alternative estimator with a lower variance. Having a lower variance means there is a higher probability of obtaining a coefficient estimate close to its true value. It also means that hypothesis tests have greater power and a lower probability of a Type II error.
2. The formulas for the standard errors usually computed for the least squares estimator are no longer correct, and hence confidence intervals and hypothesis tests that use these standard errors may be misleading.

Although the usual least squares standard errors are not the correct ones, it is possible to compute correct standard errors for the least squares estimator when the errors are autocorrelated. These standard errors are known as **HAC (heteroskedasticity and autocorrelation consistent) standard errors**, or **Newey-West standard errors**, and are analogous to the heteroskedasticity consistent standard errors introduced in Chapter 8. By using HAC standard errors with least squares, we can avoid having to specify the precise nature of the autocorrelated error model that is required for an alternative estimator with a lower variance. For the HAC standard errors to be valid, we need to assume that the autocorrelations go to zero as the time between observations increases (a condition necessary for stationarity), and we need a large sample, but it is not necessary to make a precise assumption about the autocorrelated error model.

To get a feel for how HAC standard errors are found, consider the simple regression model $y_t = \beta_1 + \beta_2 x_t + e_t$. From Appendix 8A the variance of the least squares estimator b_2 can be written as (with subscripts i and j replaced by t and s)

$$\begin{aligned}\operatorname{var}(b_2) &= \sum_t w_t^2 \operatorname{var}(e_t) + \sum_{t \neq s}\sum w_t w_s \operatorname{cov}(e_t, e_s) \\ &= \sum_t w_t^2 \operatorname{var}(e_t)\left[1 + \frac{\sum_{t\neq s}\sum w_t w_s \operatorname{cov}(e_t, e_s)}{\sum_t w_t^2 \operatorname{var}(e_t)}\right]\end{aligned} \tag{9.27}$$

where $w_t = (x_t - \bar{x})/\sum_t (x_t - \bar{x})^2$. When the errors are not correlated, $\operatorname{cov}(e_t, e_s) = 0$, and the term in square brackets is equal to one. The resulting expression $\operatorname{var}(b_2) = \sum_t w_t^2 \operatorname{var}(e_t)$ is the one used to find heteroskedasticity-consistent (HC) standard errors. When the errors are correlated, the term in square brackets is estimated to obtain HAC standard errors. If we call the quantity in square brackets g and its estimate $\hat{g}$, then the relationship between the two estimated variances is

$$\widehat{\operatorname{var}_{HAC}(b_2)} = \widehat{\operatorname{var}_{HC}(b_2)} \times \hat{g} \tag{9.28}$$

The HAC variance estimate is equal to the HC variance estimate multiplied by an extra term that depends on the serial correlation in the errors.

This explanation is a simplified one because it treats x as nonrandom. If x is random and e_t is independent of all x values, as specified in assumption TSMR2, then essentially the same argument holds. More general arguments allow for correlation between e_t and the x values, as will occur if the model contains a lagged dependent variable, and they extend the results to the multiple regression model with more than one x. However, in all cases the end result is an expression like (9.28). Several alternative estimators for g are available. They differ depending on the number of lags for which autocorrelations are estimated and on the weights placed on the autocorrelations at each lag. Because a large number of alternatives are possible, you will discover that different software packages may yield different HAC standard errors; also, different options are possible within a given software package. The message is: Don't be disturbed if you see slightly different HAC standard errors computed for the same problem.

The least squares-estimated Phillips curve $INF_t = \beta_1 + \beta_2 DU_t + e_t$ with both sets of standard errors—the incorrect least squares ones that ignore autocorrelation, and the correct HAC ones that recognize the autocorrelation—are as follows:[10]

$$
\begin{array}{llll}
\widehat{INF} = 0.7776 - 0.5279DU & & & \\
\qquad\quad (0.0658) \quad (0.2294) & & \text{(incorrect se)} & \qquad (9.29) \\
\qquad\quad (0.1030) \quad (0.3127) & & \text{(HAC se)} &
\end{array}
$$

The HAC standard errors are larger than those from least squares, implying that if we ignore the autocorrelation, we will overstate the reliability of the least squares estimates. The t and p-values for testing $H_0: \beta_2 = 0$ are

$$
\begin{array}{lll}
t = -0.5279/0.2294 = -2.301 & p = 0.0238 & \text{(from LS standard errors)} \\
t = -0.5279/0.3127 = -1.688 & p = 0.0950 & \text{(from HAC standard errors)}
\end{array}
$$

With least squares standard errors, a two-tail test, and a 5% significance level, we reject $H_0: \beta_2 = 0$. With HAC standard errors, we do not reject H_0. Thus, using incorrect standard errors can lead to misleading results. A similar conclusion can be reached by examining the 95% interval estimates for β_2 for each set of standard errors. Using $t_{(0.975,\,88)} = 1.987$, those interval estimates are $(-0.984, -0.072)$ for least squares and $(-1.149, 0.094)$ for HAC standard errors. The narrower least squares interval leads to an exaggerated conclusion about the reliability of estimation.

9.5.2 Estimating an AR(1) Error Model

Using least squares with HAC standard errors overcomes the negative consequences that autocorrelated errors have for least squares standard errors. However, it does not address the issue of finding an estimator that is better, in the sense that it has a lower variance. One way to proceed is to make an assumption about the model that generates the autocorrelated errors, and to derive an estimator compatible with this assumption. In this section we examine one such assumption. To introduce it, we return to the Lagrange multiplier test for serially correlated errors, where correlation between e_t and e_{t-1} was modeled by writing e_t as dependent on e_{t-1} through the equation

[10] The HAC standard errors were computed by EViews 7.0 using a Bartlett kernel, a Newey-West fixed bandwidth of 4, and a degrees-of-freedom adjustment.

$$e_t = \rho e_{t-1} + v_t \tag{9.30}$$

If we assume the v_t are uncorrelated random errors with zero mean and constant variances,

$$E(v_t) = 0 \qquad \mathrm{var}(v_t) = \sigma_v^2 \qquad \mathrm{cov}(v_t, v_s) = 0 \quad \text{for } t \neq s \tag{9.31}$$

then (9.30) describes a **first-order autoregressive model** or a first-order autoregressive **process** for e_t. The term **AR(1) model** is used as an abbreviation for first-order autoregressive model. It is called an **autoregressive** model because it can be viewed as a regression model where e_t depends on its lagged value, inducing autocorrelation. It is called **first-order** because the right-hand-side variable is e_t lagged **one** period.

One way to estimate a regression equation with serially correlated errors is to assume that those errors follow an AR(1) model and to develop an estimation procedure relevant for that model. Other autocorrelated error models could be assumed. In particular, one could include more lags of e_t leading to, say, an AR(2) or an AR(3) model. However, for the moment we focus on the AR(1) error model because it has been a popular one in econometrics and is a good starting point.

9.5.2a Properties of an AR(1) Error

Before turning to estimation, it is useful to examine the properties of e_t when it follows an AR(1) process. In (9.31) we made assumptions about v_t that are the same as those made about e_t in Chapters 2–7. The question now is: How do the assumptions about v_t in (9.31), and the AR(1) error model, change the properties of e_t? We make one further assumption to ensure the e_t are stationary: namely, that ρ is less than one in absolute value. That is,

$$-1 < \rho < 1 \tag{9.32}$$

In Appendix 9B, we show that the mean and variance of e_t are

$$E(e_t) = 0 \qquad \mathrm{var}(e_t) = \sigma_e^2 = \frac{\sigma_v^2}{1-\rho^2} \tag{9.33}$$

The AR(1) error e_t has a mean of zero, and a variance that depends on the variance of v_t and the magnitude of ρ. The larger the degree of autocorrelation (the closer ρ is to $+1$ or -1), the larger the variance of e_t. Also, since $\sigma_v^2/(1-\rho^2)$ is constant over time, e_t is homoskedastic.

In Appendix 9B we also discover that the covariance between two errors that are k periods apart $(e_t$ and $e_{t-k})$ is

$$\mathrm{cov}(e_t, e_{t-k}) = \frac{\rho^k \sigma_v^2}{1-\rho^2}, \quad k > 0 \tag{9.34}$$

This expression shows how the properties of the e_t differ from those assumed in Chapters 2–7. In these earlier chapters we assumed that the covariance between errors for different observations was zero. It is now nonzero because of the existence of a lagged relationship between the errors from different time periods.

It is useful to describe the *correlation* implied by the covariance in (9.34). Using the correlation formula in (9.12), we have

$$\begin{aligned}\rho_k = \mathrm{corr}(e_t, e_{t-k}) &= \frac{\mathrm{cov}(e_t, e_{t-k})}{\sqrt{\mathrm{var}(e_t)\mathrm{var}(e_{t-k})}} = \frac{\mathrm{cov}(e_t, e_{t-k})}{\mathrm{var}(e_t)} \\ &= \frac{\rho^k \sigma_v^2/(1-\rho^2)}{\sigma_v^2/(1-\rho^2)} \\ &= \rho^k\end{aligned} \tag{9.35}$$

That is, $\rho_k = \rho^k$. The correlation between two errors that are k periods apart (ρ_k) is given by ρ raised to the power k. An interpretation or definition of the unknown parameter ρ can be obtained by setting $k = 1$. Specifically,

$$\rho_1 = \mathrm{corr}(e_t, e_{t-1}) = \rho \tag{9.36}$$

Thus, ρ represents the correlation between two errors that are one period apart; it is the **first-order autocorrelation** for e, sometimes simply called the autocorrelation coefficient. Recall the concept of a correlogram that was introduced in Section 9.3. It consisted of the sequence of **sample** autocorrelations $r_1, r_2, r_3, \ldots$. The coefficient ρ is the **population** autocorrelation at lag one for a time series that can be described by an AR(1) model; r_1 is an estimate for ρ when we assume a series is AR(1).

Corresponding to the sample correlogram $r_1, r_2, r_3, \ldots,$ we can also define a population correlogram as $\rho_1, \rho_2, \rho_3, \ldots$. From (9.35), the population correlogram for an AR(1) model is $\rho, \rho^2, \rho^3, \ldots$. Because $-1 < \rho < 1$, these autocorrelations decline geometrically as the lag increases, eventually becoming negligible. Since the AR(1) error model $e_t = \rho e_{t-1} + v_t$ only contains one lag of e, you might be surprised to find that autocorrelations at lags greater than one, although declining, are still nonzero. The correlation persists because each e_t depends on all past values of the errors v_t through the equation (see Appendix 9B)

$$e_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots \tag{9.37}$$

We can relate these results to the errors from the Phillips curve. The sample correlogram for the first five lags was found to be

$$r_1 = 0.549 \qquad r_2 = 0.456 \qquad r_3 = 0.433 \qquad r_4 = 0.420 \qquad r_5 = 0.339$$

Without any assumptions about the model that generates the errors, these values are unrestricted estimates of the population autocorrelations ($\rho_1, \rho_2, \rho_3, \rho_4, \rho_5$). Now suppose the errors follow an AR(1) model where we have only one unknown parameter ρ. In this case,

$$\hat{\rho}_1 = \hat{\rho} = r_1 = 0.549$$

Imposing the structure of the AR(1) model leads to the following estimates at longer lags:

$$\begin{aligned}\hat{\rho}_2 &= \hat{\rho}^2 = (0.549)^2 = 0.301 \\ \hat{\rho}_3 &= \hat{\rho}^3 = (0.549)^3 = 0.165 \\ \hat{\rho}_4 &= \hat{\rho}^4 = (0.549)^4 = 0.091 \\ \hat{\rho}_5 &= \hat{\rho}^5 = (0.549)^5 = 0.050\end{aligned}$$

These values are considerably smaller than the unrestricted estimates of the correlogram, suggesting that the AR(1) assumption might not be adequate for the errors of the Phillips curve.

9.5.2b Nonlinear Least Squares Estimation

In this section we develop an estimator for the regression model with AR(1) errors. First, let us summarize the model and its assumptions. It is given by

$$y_t = \beta_1 + \beta_2 x_t + e_t \quad \text{with} \quad e_t = \rho e_{t-1} + v_t \tag{9.38}$$

and $-1 < \rho < 1$. Only one explanatory variable is included, to keep the discussion simple and to use the framework of the Phillips curve example. The v_t are uncorrelated random variables with mean zero and a constant variance σ_v^2 (see assumptions MR2, MR3, and MR4, stated in Section 5.1):

$$E(v_t) = 0 \qquad \text{var}(v_t) = \sigma_v^2 \qquad \text{cov}(v_t, v_s) = 0 \quad \text{for } t \neq s \tag{9.39}$$

Substituting $e_t = \rho e_{t-1} + v_t$ into $y_t = \beta_1 + \beta_2 x_t + e_t$ yields

$$y_t = \beta_1 + \beta_2 x_t + \rho e_{t-1} + v_t \tag{9.40}$$

From the regression equation the error in the previous period can be written as

$$e_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{t-1} \tag{9.41}$$

Multiplying (9.41) by ρ yields

$$\rho e_{t-1} = \rho y_{t-1} - \rho\beta_1 - \rho\beta_2 x_{t-1} \tag{9.42}$$

Substituting (9.42) into (9.40) yields

$$y_t = \beta_1(1-\rho) + \beta_2 x_t + \rho y_{t-1} - \rho\beta_2 x_{t-1} + v_t \tag{9.43}$$

What have we done? We have transformed the original model in (9.38) with the autocorrelated error term e_t into a new model given by (9.43) that has an error term v_t that is uncorrelated over time. The advantage of doing so is that we can now proceed to find estimates for (β_1, β_2, ρ) that minimize the sum of squares of uncorrelated errors $S_v = \sum_{t=2}^{T} v_t^2$. Minimizing the sum of squares of the correlated errors $S_e = \sum_{t=1}^{T} e_t^2$ yields the least squares estimator that is not best and whose standard errors are not correct. However, minimizing the sum of squares of uncorrelated errors, S_v, yields an estimator that is best and whose standard errors are correct (in large samples). Note that this result is in line with earlier practice in the book. The least squares estimator used in Chapters 2–7 minimizes a sum of squares of uncorrelated errors.

There are, however, two important distinctive features about the transformed model in (9.43). To appreciate the first, note that the coefficient of x_{t-1} is equal to $-\rho\beta_2$ which is the negative product of ρ (the coefficient of y_{t-1}) and β_2 (the coefficient of x_t). This fact means that although (9.43) is a linear function of the variables x_t, y_{t-1} and x_{t-1}, it is not a linear function of the parameters (β_1, β_2, ρ). The usual linear least squares formulas cannot be obtained by using calculus to find the values of (β_1, β_2, ρ) that minimize S_v. Nevertheless, computer software can be used to find the estimates numerically. Numerical methods use a systematic procedure for trying a sequence of alternative parameter values until those which minimize the sum of squares function are found. Because these estimates are not computed from a linear formula but still minimize a sum of squares function, they are called

nonlinear least squares estimates. Estimates obtained in this way have the usual desirable properties in large samples under assumptions TSMR2A and TSMR3–5.

The second distinguishing feature about the model in (9.43) is that it contains the lagged dependent variable y_{t-1} as well as x_t and x_{t-1}, the current and lagged values of the explanatory variable. For this reason, the summation $S_v = \sum_{t=2}^{T} v_t^2$ begins at $t = 2$.

In the last section, we cast some doubt on whether the AR(1) error model was an appropriate one for capturing the residual autocorrelations in the Phillips curve example. Nevertheless, we will estimate the Phillips curve assuming AR(1) errors; later, we investigate whether a better model can be found. In this context, (9.43) becomes

$$INF_t = \beta_1(1-\rho) + \beta_2 DU_t + \rho INF_{t-1} - \rho\beta_2 DU_{t-1} + v_t \tag{9.44}$$

Applying nonlinear least squares and presenting the estimates in terms of the original untransformed model, we have

$$\begin{array}{llll} \widehat{INF} = 0.7609 - 0.6944DU & & e_t = 0.557e_{t-1} + v_t \\ (\text{se}) \quad (0.1245) \;\; (0.2479) & & \qquad (0.090) \end{array} \tag{9.45}$$

Comparing these estimates with those from least squares $(b_1 = 0.7776,\ b_2 = -0.5279)$, we find that the estimate for β_1 is of similar magnitude, but that that for β_2 is a larger negative value, suggesting a greater impact of unemployment on inflation. The standard error $\text{se}(\hat{\beta}_2) = 0.2479$ is smaller than the corresponding HAC least squares standard error $[\text{se}(b_2) = 0.3127]$, suggesting a more reliable estimate, but the standard error for $\hat{\beta}_1$ is unexpectedly larger, something that we do not expect since we have used an estimation procedure with a lower variance. It must be kept in mind, however, that standard errors are themselves estimates of true underlying standard deviations. The estimate $\hat{\rho} = 0.557$ is similar but not exactly the same as the estimate $r_1 = 0.549$ obtained from the correlation between least squares residuals that are one quarter apart.

9.5.2c Generalized Least Squares Estimation

In Chapter 8 we discovered that the problem of heteroskedasticity could be overcome by using an estimation procedure known as generalized least squares, and that a convenient way to obtain generalized least squares estimates is to first transform the model so that it has a new uncorrelated homoskedastic error term, and to then apply least squares to the transformed model. This same kind of approach can be pursued when the errors follow an AR(1) model. Indeed, it can be shown that nonlinear least squares estimation of (9.43) is equivalent to using an iterative generalized least squares estimator called the Cochrane-Orcutt procedure. Details are provided in Appendix 9C.

9.5.3 Estimating a More General Model

The results for the Phillips curve example presented in (9.45) came from estimating the AR (1) error model written as the transformed model

$$y_t = \beta_1(1-\rho) + \rho y_{t-1} + \beta_2 x_t - \rho\beta_2 x_{t-1} + v_t \tag{9.46}$$

Suppose now that we consider the model

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t \tag{9.47}$$

How do (9.46) and (9.47) differ? What characteristics do they have in common? The first thing to notice is that they contain the same variables; in both cases y_t depends on x_t, x_{t-1} and y_{t-1}. There is a difference in the number of parameters, however. In (9.46) there are three unknown parameters, β_1, β_2, and ρ. In (9.47) there are four unknown parameters, δ, δ_0, δ_1, and θ_1. Also, the notation in (9.47) is new; we have used the symbols δ (delta) and θ (theta). The intercept is denoted by δ, the coefficients of x and its lag are denoted by subscripted δ's, and the coefficient of the lagged dependent variable y_{t-1} is given by a subscripted θ. This new notation will prove to be convenient in Section 9.6, where we discuss a general class of **autoregressive distributed lag (ARDL) models**. Equation (9.47) is a member of this class.

To establish the relationship between (9.46) and (9.47), note that (9.47) is the same as (9.46) if we set

$$\delta = \beta_1(1-\rho) \qquad \delta_0 = \beta_2 \qquad \delta_1 = -\rho\beta_2 \qquad \theta_1 = \rho \tag{9.48}$$

From these relationships, it can be seen that (9.46) is a restricted version of (9.47) with the restriction $\delta_1 = -\theta_1\delta_0$ imposed. The restriction reduces the number of parameters from four to three and makes (9.47) equivalent to the AR(1) error model.

These observations raise a number of questions. Instead of estimating the AR(1) error model, would it be better to estimate the more general model in (9.47)? What technique should be used for estimating (9.47)? Is it possible to estimate (9.47) and then test the validity of the AR(1) error model by testing a null hypothesis $H_0 : \delta_1 = -\theta_1\delta_0$?

Considering estimation first, we note that (9.47) can be estimated by linear least squares providing that the v_t satisfy the usual assumptions required for least squares estimation—namely, that they have zero mean and constant variance and are uncorrelated. The presence of the lagged dependent variable y_{t-1} means that a large sample is required for the desirable properties of the least squares estimator to hold, but the least squares procedure is still valid providing that assumption TSMR2A holds. It is important for the v_t to be uncorrelated. If they are correlated, assumption TSMR2A will be violated, and the least squares estimator will be biased, even in large samples.

In the introduction to this chapter we observed that dynamic characteristics of time-series relationships can occur through lags in the dependent variable, lags in the explanatory variables, or lags in the error term. In this section we modeled a lag in the error term with an AR(1) process and showed that such a model is equivalent to (9.46), which, in turn, is a special case of (9.47). Notice that (9.46) and (9.47) do not have lagged error terms, but they do have a lagged dependent variable and a lagged explanatory variable. Thus, the dynamic features of a model implied by an AR(1) error can be captured by using instead a model with a lagged y and a lagged x. This observation raises issues about a general modeling strategy for dynamic economic relationships. Instead of explicitly modeling lags through an autocorrelated error, we may be able to capture the same dynamic effects by adding lagged variables y_{t-1} and x_{t-1} to the original linear equation.

Is it possible to test $H_0 : \delta_1 = -\theta_1\delta_0$ and hence decide whether the AR(1) model is a reasonable restricted version of (9.47) or whether the more general model in (9.47) would be preferable? The answer is yes: the test is similar to, but more complicated than, those considered in Chapter 5. Complications occur because the hypothesis involves an equation that is nonlinear in the parameters, and the delta method (see Appendix 5B) is needed to compute the standard error of products such as $\hat{\theta}_1\hat{\delta}_0$. Nevertheless, the test, known as a Wald test, can be performed using most software.

Applying the least squares estimator to (9.47) using the data for the Phillips curve example yields

$$\widehat{INF}_t = 0.3336 + 0.5593INF_{t-1} - 0.6882DU_t + 0.3200DU_{t-1}$$
$$(\text{se}) \quad (0.0899) \quad (0.0908) \qquad (0.2575) \qquad (0.2499) \tag{9.49}$$

How do these results compare with those from the more restrictive AR(1) error model? Most of them turn out to be very similar. The equivalent estimates from the AR(1) error model are found by substituting the estimates in (9.45) into the expressions in (9.48). We find

$$\hat{\delta} = \hat{\beta}_1(1-\hat{\rho}) = 0.7609 \times (1 - 0.5574) = 0.3368 \quad \text{which is similar to } 0.3336$$
$$\hat{\theta}_1 = \hat{\rho} = 0.5574 \quad \text{which is similar to } 0.5593$$
$$\hat{\delta}_0 = \hat{\beta}_2 = -0.6944 \quad \text{which is similar to } -0.6882$$
$$\hat{\delta}_1 = -\hat{\rho}\hat{\beta}_2 = -0.5574 \times (-0.6944) = 0.3871 \quad \text{which differs a little from } 0.3200$$

The closeness of these values and the relatively large standard error on the coefficient of DU_{t-1} suggest that a test of the restriction $H_0: \delta_1 = -\theta_1\delta_0$ would not be rejected. More formally, using a Wald chi-square test yields a value of $\chi^2_{(1)} = 0.112$ with a corresponding p-value $= 0.738$. On the basis of this test, we conclude that the AR(1) error model is not too restrictive.

Specification and estimation of the more general model does have some advantages, however. It makes the dependence of y_t on its lag and that of x more explicit, and it can often provide a useful economic interpretation. The original economic model for the Phillips curve was

$$INF_t = INF_t^E - \gamma(U_t - U_{t-1}) \tag{9.50}$$

Comparing this model with the estimated one in (9.50), an estimate of the model for inflationary expectations is $INF_t^E = 0.3336 + 0.5593INF_{t-1}$; expectations for inflation in the current quarter are 0.33% plus 0.56 times last quarter's inflation rate. The effect of unemployment in (9.49) is $-0.6882(U_t - U_{t-1}) + 0.3200(U_{t-1} - U_{t-2})$, which is dynamically more complex than the original specification of $-\gamma(U_t - U_{t-1})$. Note, however, that the coefficient of DU_{t-1} is not significantly different from zero in (9.49). If DU_{t-1} is excluded from the equation, then the unemployment effect is consistent with the original equation. Re-estimation of the model after omitting DU_{t-1} yields

$$\widehat{INF}_t = 0.3548 + 0.5282INF_{t-1} - 0.4909DU_t$$
$$(\text{se}) \quad (0.0876) \quad (0.0851) \qquad (0.1921) \tag{9.51}$$

In this model inflationary expectations are given by $INF_t^E = 0.3548 + 0.5282INF_{t-1}$ and a 1% rise in the unemployment rate leads to an approximate 0.5% fall in the inflation rate.

9.5.4 Summary of Section 9.5 and Looking Ahead

In Section 9.5 we have described three ways of overcoming the effect of serially correlated errors:

1. Estimate the model using least squares with HAC standard errors.
2. Use nonlinear least squares to estimate the model with a lagged x, a lagged y, and the restriction implied by an AR(1) error specification.

3. Use least squares to estimate the model with a lagged x and a lagged y, but without the restriction implied by an AR(1) error specification.

Using least squares with HAC standard errors is preferred if one does not wish to bother transforming the model to one with relevant lagged variables that have the effect of eliminating the serial correlation in the errors. The nonlinear model with the AR(1) error restriction is appropriate if the error has serial correlation that takes the form of an AR(1) process. For many years it was the most common method for correcting for autocorrelated errors. However, the third method—including appropriate lags of y and x without the AR(1) error restriction—is now generally preferred by applied econometricians. It is less restrictive than the AR(1) error model, the model with lags frequently has a useful economic interpretation, and it can be used to correct for more general forms of serially correlated errors than the AR(1) error model.

This last statement raises some unanswered questions. While including one lag of y and one lag of x will correct for serially correlated errors if they follow an AR(1) model, it might not solve the problem if the form of serial correlation is more complex. How do we check whether some serial correlation still remains? If we include a lagged y and a lagged x and the errors are still serially correlated, how do we proceed? Checking for serial correlation proceeds along the same lines as we have described in Section 9.4. We apply the same tests to the errors from the new model with lags. Also, if we have doubts about whether the errors in the new model are correlated, we can use HAC standard errors with this model. Alternatively, including more lags of y on the right side of the equation can have the effect of eliminating any remaining serial correlation in the errors. Models with a general number of lags of y and x are called autoregressive distributed lag models; we consider them in the next section.

9.6 Autoregressive Distributed Lag Models

An autoregressive distributed lag (ARDL) model is one that contains both lagged x_t's and lagged y_t's. In its general form, with p lags of y and q lags of x, an ARDL(p, q) model can be written as

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + v_t \tag{9.52}$$

The AR component of the name ARDL comes from the regression of y on lagged values of itself; the DL component comes from the distributed lag effect of the lagged x's. Two examples that we have encountered so far in (9.50) and (9.51) are

$$\text{ARDL(1,1): } \widehat{INF}_t = 0.3336 + 0.5593 INF_{t-1} - 0.6882 DU_t + 0.3200 DU_{t-1}$$

$$\text{ARDL(1,0): } \widehat{INF}_t = 0.3548 + 0.5282 INF_{t-1} - 0.4909 DU_t$$

The ARDL model has several advantages. It captures dynamic effects from lagged x's and lagged y's, and by including a sufficient number of lags of y and x, we can eliminate serial correlation in the errors. Moreover, an ARDL model can be transformed into one with only lagged x's which go back into the infinite past:

$$\begin{aligned} y_t &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots + e_t \\ &= \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + e_t \end{aligned} \tag{9.53}$$

Because it does not have a finite cut off point, this model is called an **infinite distributed lag model**. It contrasts with the finite distributed lag model we studied in Section 9.2, where the effect of the lagged x's was assumed to cut off to zero after q lags. Like before, the parameter β_s is the distributed lag weight or the s-period delay multiplier showing the effect of a change in x_t on y_{t+s}. The total or long-run multiplier showing the long-run effect of a sustained change in x_t is $\sum_{s=0}^{\infty}\beta_s$. For the transformation from (9.52) to (9.53) to be valid, the effect of a change must gradually die out. Thus, the values of β_s for large s will be small and decreasing, a property that is necessary for the infinite sum $\sum_{s=0}^{\infty}\beta_s$ to be finite. Estimates for the lag weights β_s can be found from estimates of the θ_k's and the δ_j's in (9.52), with the precise relationship between them depending on the values for p and q. This relationship is explored in Section 9.8.

The two main uses of ARDL models are for forecasting and multiplier analysis. Both are useful policy tools. We consider them in Sections 9.7 and 9.8, respectively. For the remainder of this section we consider *estimation* of (9.52). Because estimation is straightforward—least squares is an appropriate estimation technique under assumptions TSMR1, TSMR2A, and TSMR3–5—the main concern for estimation is choice of the lag lengths p and q.

There are a number of different criteria for choosing p and q. Because they all do not necessarily lead to the same choice, there is a degree of subjective judgment that must be used. Four possible criteria are

1. Has serial correlation in the errors been eliminated? If not, then least squares will be biased in small and large samples. It is important to include sufficient lags, especially of y, to ensure that serial correlation does not remain. It can be checked using the correlogram or Lagrange multiplier tests.
2. Are the signs and magnitudes of the estimates consistent with our expectations from economic theory? Estimates which are poor in this sense may be a consequence of poor choices for p and q, but they could also be symptomatic of a more general modeling problem.
3. Are the estimates significantly different from zero, particularly those at the longest lags?
4. What values for p and q minimize information criteria such as the AIC and SC? Information criteria were first considered in Chapter 6. In the context of the ARDL model they involve choosing p and q to minimize the sum of squared errors (SSE) subject to a penalty that increases as the number of parameters increases. Increasing lag lengths increases the number of parameters, and, providing we use the same number of observations in each case,[11] it reduces the sum of squared errors; penalty terms are included with a view to capturing the essential lag effects without introducing an excessive number of parameters. The **Akaike information criterion** (**AIC**) is given by[12]

$$\mathrm{AIC} = \ln\left(\frac{SSE}{T}\right) + \frac{2K}{T} \tag{9.54}$$

[11] Care must be taken to use the same number of observations. Unless special provision is made, the number of observations used will typically decline as the lag length increases.

[12] You will find slight but nonessential variations in the definitions of AIC and SC. For example, to get the values computed by EViews 7.0 you need to add $[1+\ln(2\pi)]$ to the expressions in (9.54) and (9.55). Adding or subtracting a constant does not change the lag length that minimizes AIC or SC.

where $K = p + q + 2$ is the number of coefficients that are estimated. The **Schwarz criterion** (**SC**), also known as the **Bayes information criterion** (**BIC**), is given by

$$\text{SC} = \ln\left(\frac{SSE}{T}\right) + \frac{K\ln(T)}{T} \tag{9.55}$$

Because $K\ln(T)/T > 2K/T$ for $T \geq 8$, the SC penalizes additional lags more heavily than does the AIC.

We now apply the above criteria to our two examples—the Phillips curve and the equation for Okun's law—to see if we can improve on our earlier specifications.

9.6.1 The Phillips Curve

Our starting point for the Phillips curve is the previously estimated ARDL(1,0) model

$$\begin{aligned} \widehat{INF}_t &= 0.3548 + 0.5282INF_{t-1} - 0.4909DU_t, \quad \text{obs} = 90 \\ (\text{se}) \quad & \quad (0.0876) \quad (0.0851) \qquad\quad (0.1921) \end{aligned} \tag{9.56}$$

We choose this model in preference to the ARDL(1,1) model because the coefficient of DU_{t-1} was not significantly different from zero. Also, to help avoid confusion that may arise because we are considering models with differing numbers of lags, we have indicated that 90 observations were used for estimation.

Checking first to see whether the errors from (9.56) are serially correlated, we obtain the correlogram for its residuals presented in Figure 9.9. Since these autocorrelations are not significantly different from zero, they provide no evidence of serial correlation. However, a further check using Lagrange multiplier tests provides conflicting evidence. Table 9.3 contains the p-values for the $LM = T \times R^2$ version of the LM test (with pre-sample errors set equal to zero) for autocorrelation of orders one to five. Using a 5% significance level, tests for orders one, four, and five reject a null hypothesis of no autocorrelation.

Taken together, these test results provide some evidence, but not overwhelming evidence, that serial correlation in the errors still exists; one lag of the dependent variable

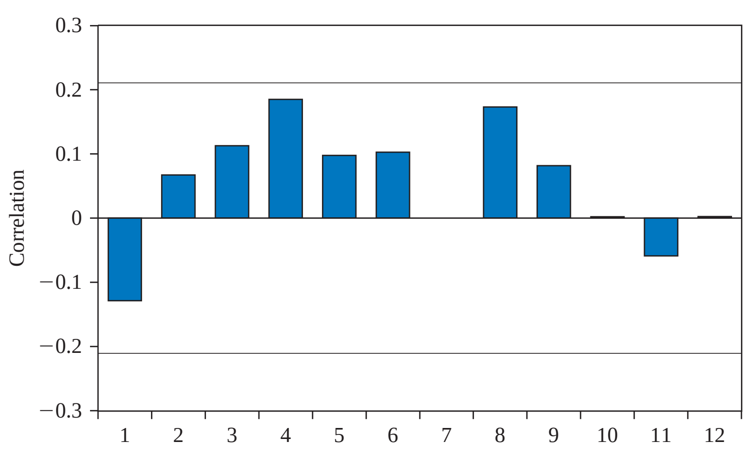

FIGURE 9.9 Correlogram for residuals from Phillips curve ARDL(1,0) model.

Table 9.3 ***p*-values for *LM* Test for Autocorrelation**

Lag	p-value
1	0.0421
2	0.0772
3	0.1563
4	0.0486
5	0.0287

INF has not been sufficient to eliminate the autocorrelation. When additional lags of both *INF* and *DU* are tried, we find

1. Coefficients of extra lags of *DU* are never significantly different from zero at a 5% level of significance.
2. For $p = 2$, $q = 0$, the coefficients of INF_{t-1} and INF_{t-2} are significantly different from zero at a 5% significance level; for $p = 3$, $q = 0$, the coefficients of INF_{t-1} and INF_{t-3} are significant; and for $p = 4$, $q = 0$, the coefficients of INF_{t-1} and INF_{t-4} are significant. Coefficients of lags greater than 4 ($p \geq 5$) were not significant. Moreover, for $p = 2$ and $p = 3$ the *LM* test continued to suggest serial correlation in the errors. For $p = 4$ no correlation remained.

Thus, if we use significance of coefficients and elimination of serial correlation in the errors as our criteria for selecting lag lengths, our choice is the ARDL(4,0) model

$$\begin{aligned}\widehat{INF}_t &= \underset{(0.0983)}{0.1001} + \underset{(0.1016)}{0.2354}INF_{t-1} + \underset{(0.1038)}{0.1213}INF_{t-2} + \underset{(0.1050)}{0.1677}INF_{t-3} \\ \text{(se)} \quad & \quad + \underset{(0.1014)}{0.2819}INF_{t-4} - \underset{(0.1885)}{0.7902}DU_t \qquad \text{obs} = 87\end{aligned} \tag{9.57}$$

In this model inflationary expectations are given by

$$INF_t^E = 0.1001 + 0.2354INF_{t-1} + 0.1213INF_{t-2} + 0.1677INF_{t-3} + 0.2819INF_{t-4}$$

A relatively large weight is given to actual inflation in the corresponding quarter of the previous year ($t-4$). The effect of unemployment on inflation is larger in this model. A 1% rise in unemployment reduces inflation by approximately 0.8%.

Table 9.4 contains the AIC and SC values for $p = 1$ to 6 and $q = 0, 1$. To compute these values, 85 observations were used for all cases, with the starting quarter being 1988, quarter 3. The values that minimize both the AIC and the SC (the largest negative values) are $p = 4$ and $q = 0$, supporting the choice of the ARDL(4,0) model given in (9.57).[13]

[13] Since the coefficients of INF_{t-2} and INF_{t-3} are not significantly different from zero, we could also consider dropping one or both of these terms from the equation but retaining INF_{t-4}. If one follows this strategy, the model that minimizes the AIC and the SC omits INF_{t-2} but keeps INF_{t-1}, INF_{t-3}, and INF_{t-4}.

Table 9.4 **AIC and SC Values for Phillips Curve ARDL Models**

p	q	AIC	SC	p	q	AIC	SC
1	0	−1.247	−1.160	1	1	−1.242	−1.128
2	0	−1.290	−1.176	2	1	−1.286	−1.142
3	0	−1.335	−1.192	3	1	−1.323	−1.151
4	0	−1.402	−1.230	4	1	−1.380	−1.178
5	0	−1.396	−1.195	5	1	−1.373	−1.143
6	0	−1.378	−1.148	6	1	−1.354	−1.096

9.6.2 Okun's Law

In Section 9.2.1 we estimated an equation for Okun's law. It was given by the following finite distributed lag model where the change in unemployment (DU) was related to GDP growth (G) and its lags

$$\begin{array}{lllll} \widehat{DU}_t = & 0.5836 - & 0.2020G_t - & 0.1653G_{t-1} - & 0.0700G_{t-2} \\ \text{(se)} & (0.0472) & (0.0324) & (0.0335) & (0.0331) \end{array} \quad \text{obs} = 96 \qquad (9.58)$$

In the more general ARDL context, this equation is an ARDL(0,2) model. It has no lags of DU and two lags of G. We now ask whether we can improve on this model. Does it suffer from serially correlated errors? If we include lagged values of DU, do those lags have coefficients that are significantly different from zero?

The correlogram for the residuals from (9.58) is displayed in Figure 9.10. It shows a significant autocorrelation at lag one, with the remaining autocorrelations being insignificant. This correlation is confirmed by the LM test whose p-value is 0.0004 for a test with lag order one and pre-sample residual set to zero. When DU_{t-1} is included with a view to eliminating the serial correlation, we find that its coefficient is significantly different from zero, but that for G_{t-2} becomes insignificant. The estimated equation is

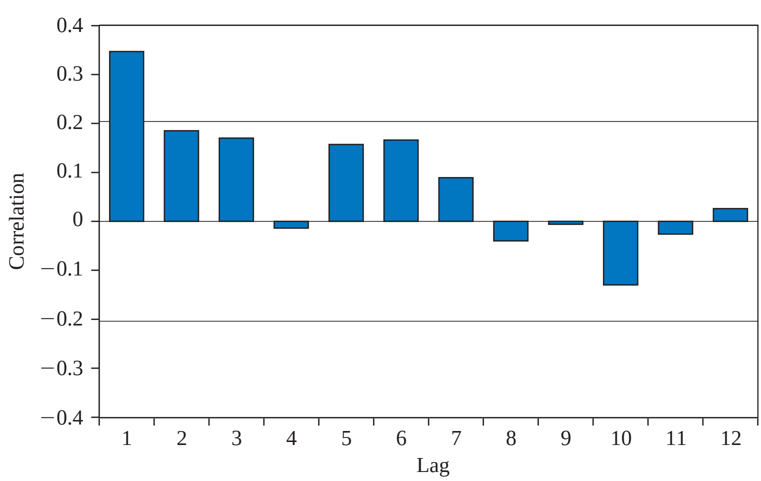

FIGURE 9.10 Correlogram for residuals from Okun's law ARDL(0,2) model.

Table 9.5 AIC and SC Values for Okun's Law ARDL Models

(p, q)	AIC	SC	(p, q)	AIC	SC	(p, q)	AIC	SC
(0,1)	−3.436	−3.356	(1,1)	−3.588	−3.480	(2,1)	−3.569	−3.435
(0,2)	−3.463	−3.356	(1,2)	−3.568	−3.433	(2,2)	−3.548	−3.387
(0,3)	−3.442	−3.308	(1,3)	−3.561	−3.400	(2,3)	−3.549	−3.361

$$\begin{array}{llll} \widehat{DU}_t = 0.3780 + 0.3501DU_{t-1} - 0.1841G_t - 0.0992G_{t-1} \\ \text{(se)} \quad (0.0578) \quad (0.0846) \qquad\quad (0.0307) \qquad (0.0368) \end{array} \qquad obs = 96 \tag{9.59}$$

There is no evidence that the residuals from (9.59) are autocorrelated. Both the correlogram and *LM* test failed to reject null hypotheses of zero autocorrelations. Furthermore, when extra lags of *DU* and *G* are added to (9.59), their coefficients are not significantly different from zero at a 5% significance level. Thus, we are led to conclude that the ARDL(1,1) model in (9.59) is a suitable one for modeling the relationship between *DU* and *G*. As a final check we can examine what values of p and q minimize the AIC and SC criteria. Table 9.5 contains the AIC and SC values for possibly relevant lags. They support our choice of the ARDL(1,1) model; both criteria are at a minimum when $p = q = 1$. They were calculated using 95 observations with a starting period of 1986, quarter 1.

We examine how this model can be used for forecasting in Section 9.7. In Section 9.8 we derive multipliers showing the effect of a change in the growth rate of GDP on changes in the unemployment rate.

9.6.3 Autoregressive Models

The ARDL models in the previous section had an autoregressive component (lagged values of the dependent variable y) and a distributed lag component (an explanatory variable x, and its lags). One special case of an ARDL model is the finite distributed lag model that has no autoregressive component ($p = 0$). We studied this model in Section 9.2. It is also possible to have a pure autoregressive (AR) model with only lagged values of the dependent variable as the right-hand-side variables, and no distributed-lag component. Specifically, an autoregressive model of order p, denoted AR(p), is given by

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + v_t \tag{9.60}$$

In this model the current value of a variable y_t depends on its values in the last p periods and on a random error that is assumed to have a zero mean and a constant variance, and to be uncorrelated over time. The order of the model p is equal to the largest lag of y on the right side of the equation. Notice that there are no explanatory variables in (9.60). The value of y_t depends only on a history of its past values and no x's.

In Section 9.5.2 we were concerned with an AR(1) error model $e_t = \rho e_{t-1} + v_t$ and its implications for estimating β_1 and β_2 in the regression model $y_t = \beta_1 + \beta_2 x_t + e_t$. What is now evident is that the AR class of models has wider applicability than its use for modeling dynamic error terms. It is also used for modeling observed values of a time series y_t. The main use of AR models is for forecasting. Multiplier analysis, where the effect on y of a change in x is traced through time, is no longer possible in the absence of an x. When (9.60) is used for forecasting, we are using the current and past values of a variable to forecast its future value. The model relies on correlations between values of the variable over time to help produce a forecast.

As an example, consider the data on growth of U.S. GDP from quarter 2, 1985 to quarter 3, 2009, stored in the file *okun.dat*. This series was graphed in Figure 9.4(b), and its correlogram is displayed in Figure 9.6. Go back and look at the correlogram. The correlation between G_t and G_{t-1} (observations that are one quarter apart) is 0.494, and the correlation between G_t and G_{t-2} (observations that are two quarters apart) is 0.411. Both are significantly different from zero. How many lags are needed—what value of p is required—for an AR model to capture these correlations? Recall from Section 9.5.2a that the population autocorrelations from an AR(1) model are given by $\rho_k = \rho^k$ where k is the order of the lag. In particular, $\rho_1 = \rho$ and $\rho_2 = \rho_1^2$. Thus, for an AR(1) model to be adequate for G, we would expect $r_2 = 0.411$ to be approximately equal to the square of $r_1 = 0.494$. However, $r_1^2 = (0.494)^2 = 0.244$, which is quite a bit smaller than 0.411. It is likely that the extra correlation will be captured by an AR(2) model, and so we begin with the estimated model

$$\begin{aligned} \widehat{G}_t &= \underset{(0.1433)}{0.4657} + \underset{(0.1000)}{0.3770}G_{t-1} + \underset{(0.1029)}{0.2462}G_{t-2} \\ (\text{se}) & \end{aligned} \qquad \text{obs} = 96 \qquad (9.61)$$

The coefficient of G_{t-2} is significantly different from zero at a 5% level, suggesting we do need at least two lags of G. To check whether two lags are adequate we follow the same steps that were used for selecting the lag orders in an ARDL model. The possibility of serially correlated errors is assessed using the correlogram of the residuals and *LM* tests. Extra lags of G are added to see if their coefficients are significantly different from zero. And we can check what value of p minimizes the AIC and SC criteria.

The correlogram for the residuals is displayed in Figure 9.11. With the exception of a slightly significant autocorrelation at lag 12, all autocorrelations are not significantly different from zero at a 5% level. Since the correlations at lags 1 to 11 are insignificant, we are inclined not to react strongly to the result at lag 12. Also, *LM* tests using various orders of lags did not reveal any residual autocorrelation, and when extra lags of G were added, their coefficients were not significantly different from zero. All these results point towards the AR(2) model in (9.61) as a suitable model.

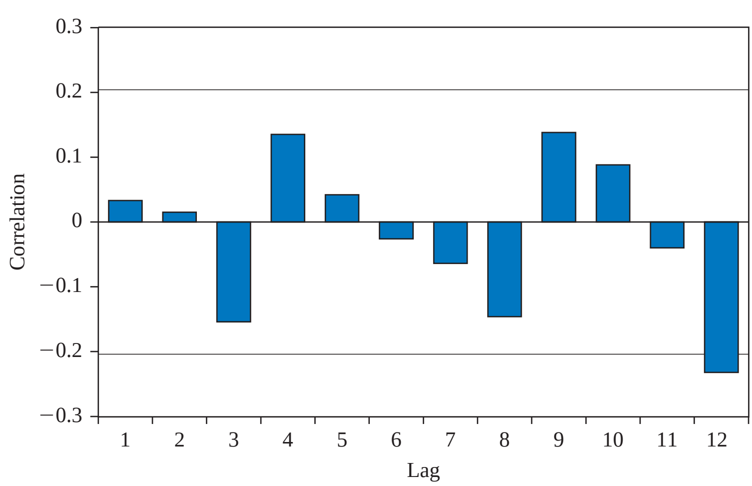

FIGURE 9.11 Correlogram for residuals from AR(2) model for GDP growth.

Table 9.6 AIC and SC Values for AR Model of Growth in U.S. GDP

Order (p)	1	2	3	4	5
AIC	−1.094	−1.131	−1.124	−1.133	−1.112
SC	−1.039	−1.049	−1.015	−0.997	−0.948

The AIC and SC values for lags up to five using a starting date of 1986, quarter 3 and 93 observations are given in Table 9.6. In this case the lag length that minimizes the AIC is different from that which minimizes the SC. Specifically, the SC suggests that we choose $p = 2$ and thus supports our earlier choice of an AR(2) model, whereas the AIC suggests a longer lag length of $p = 4$. The SC imposes a heavier penalty for the longer lag. You will find instances like this where different strategies for model choice lead to different outcomes, making some subjective judgment necessary. We will retain the AR(2) model and move to the next section where we show how to use it for forecasting.

9.7 Forecasting

Forecasting values of economic variables is a major activity for many institutions, including firms, banks, governments, and individuals. Accurate forecasts are important for decision-making on government economic policy, investment strategies, the supply of goods to retailers, and a multitude of other things that affect our everyday lives. Because of its importance, you will find that there are whole books and courses that are devoted to the various aspects of forecasting—methods and models for forecasting, ways of evaluating forecasts and their reliability, and practical examples. In this section we consider forecasting using three different models, an AR model, an ARDL model, and an exponential smoothing model. Our focus is on short-term forecasting, typically up to three periods into the future.

9.7.1 Forecasting with an AR Model

Suppose that it is the third quarter in 2009, you have estimated the AR(2) model in (9.61) using observations on growth in U.S. GDP up to and including that for 2009Q3, and you would like to forecast GDP growth for the next three quarters: 2009Q4, 2010Q1, and 2010Q2. How do we use the AR(2) model to give these forecasts? How do we calculate standard errors for our forecasts? What about forecast intervals?

We begin by writing the AR(2) model in terms of its unknown coefficients

$$G_t = \delta + \theta_1 G_{t-1} + \theta_2 G_{t-2} + v_t \tag{9.62}$$

Denoting the last sample observation as G_T, our task is to forecast G_{T+1}, G_{T+2}, and G_{T+3}. Using (9.62), we can obtain the equation that generates G_{T+1} by changing the time subscripts. The required equation is

$$G_{T+1} = \delta + \theta_1 G_T + \theta_2 G_{T-1} + v_{T+1}$$

Recognizing that the growth values for the two most recent quarters are $G_T = G_{2009Q3} = 0.8$, and $G_{T-1} = G_{2009Q2} = -0.2$, the forecast of $G_{T+1} = G_{2009Q4}$ obtained from the estimated equation in (9.61) is[14]

[14] We carry the coefficient estimates to five decimal places to avoid rounding error.

$$\begin{aligned}\hat{G}_{T+1} &= \hat{\delta} + \hat{\theta}_1 G_T + \hat{\theta}_2 G_{T-1} \\ &= 0.46573 + 0.37700 \times 0.8 + 0.24624 \times (-0.2) \\ &= 0.7181\end{aligned} \tag{9.63}$$

Moving to the forecast for two quarters ahead, G_{2010Q1}, we have

$$\begin{aligned}\hat{G}_{T+2} &= \hat{\delta} + \hat{\theta}_1 \hat{G}_{T+1} + \hat{\theta}_2 G_T \\ &= 0.46573 + 0.37700 \times 0.71808 + 0.24624 \times 0.8 \\ &= 0.9334\end{aligned} \tag{9.64}$$

There is an important difference in the way the forecasts $\hat{G}_{T+1}$ and $\hat{G}_{T+2}$ are obtained. It is possible to calculate $\hat{G}_{T+1}$ using only past observations on y. However, G_{T+2} depends on G_{T+1}, which is unobserved at time T. To overcome this problem, we replace G_{T+1} by its forecast $\hat{G}_{T+1}$ on the right side of (9.64). For forecasting G_{T+3}, the forecasts for both G_{T+2} and G_{T+1} are needed on the right side of the equation. Specifically,

$$\begin{aligned}\hat{G}_{T+3} &= \hat{\delta} + \hat{\theta}_1 \hat{G}_{T+2} + \hat{\theta}_2 \hat{G}_{T+1} \\ &= 0.46573 + 0.37700 \times 0.93343 + 0.24624 \times 0.71808 \\ &= 0.9945\end{aligned} \tag{9.65}$$

The forecast growth rates for 2009Q4, 2010Q1, and 2010Q2 are approximately 0.72%, 0.93%, and 0.99%, respectively.

We are typically interested not just in point forecasts, but also in interval forecasts that give a likely range in which a future value could fall and that indicate the reliability of a point forecast. A 95% interval forecast for j periods into the future is given by $\hat{G}_{T+j} \pm t_{(0.975,\, df)}\hat{\sigma}_j$ where $\hat{\sigma}_j$ is the standard error of the forecast error and df is the number of degrees of freedom in the estimation of the AR model ($df = 93$ in our example). To get the standard errors, note that the first forecast error, occurring at time $T+1$, is

$$u_1 = G_{T+1} - \hat{G}_{T+1} = (\delta - \hat{\delta}) + (\theta_1 - \hat{\theta}_1)G_T + (\theta_2 - \hat{\theta}_2)G_{T-1} + v_{T+1}$$

The difference between the forecast $\hat{G}_{T+1}$ and the corresponding realized value G_{T+1} depends on the differences between the actual coefficients and the estimated coefficients and on the value of the unpredictable random error v_{T+1}. A similar situation arose in Chapters 4 and 6 when we were forecasting using the regression model. What we are going to do differently now is to ignore the error from estimating the coefficients. It is common to do so because the variance of the random error is usually large relative to the variances of the estimated coefficients, and the resulting variance estimator retains the property of consistency. This means we can write the forecast error for one quarter ahead as

$$u_1 = v_{T+1} \tag{9.66}$$

For two quarters ahead, the forecast error gets more complicated because we have to allow for not only v_{T+2} but also for the error that occurs from using $\hat{G}_{T+1}$ instead of G_{T+1} on the right side of (9.64). Thus, the forecast error for two periods ahead is

$$u_2 = \theta_1(G_{T+1} - \hat{G}_{T+1}) + v_{T+2} = \theta_1 u_1 + v_{T+2} = \theta_1 v_{T+1} + v_{T+2} \tag{9.67}$$

Table 9.7 Forecasts and Forecast Intervals for GDP Growth

Quarter	Forecast $\hat{G}_{T+j}$	Standard Error of Forecast Error $(\hat{\sigma}_j)$	Forecast Interval $(\hat{G}_{T+j} \pm 1.9858 \times \hat{\sigma}_j)$
2009Q4 $(j=1)$	0.71808	0.55269	$(-0.379, 1.816)$
2010Q1 $(j=2)$	0.93343	0.59066	$(-0.239, 2.106)$
2010Q2 $(j=3)$	0.99445	0.62845	$(-0.254, 2.242)$

For three periods ahead the error can be shown to be

$$u_3 = \theta_1 u_2 + \theta_2 u_1 + v_{T+3} = (\theta_1^2 + \theta_2)v_{T+1} + \theta_1 v_{T+2} + v_{T+3} \tag{9.68}$$

Expressing the forecast errors in terms of the v_t's is convenient for deriving expressions for the forecast error variances. Because the v_t's are uncorrelated with constant variance σ_v^2, (9.66), (9.67), and (9.68) can be used to show that

$$\begin{aligned}
\sigma_1^2 &= \mathrm{var}(u_1) = \sigma_v^2 \\
\sigma_2^2 &= \mathrm{var}(u_2) = \sigma_v^2\left(1 + \theta_1^2\right) \\
\sigma_3^2 &= \mathrm{var}(u_3) = \sigma_v^2\left(\left(\theta_1^2 + \theta_2\right)^2 + \theta_1^2 + 1\right)
\end{aligned}$$

The standard errors of the forecast errors are obtained by replacing the unknown parameters in the above expressions by their estimates $\left(\hat{\theta}_1 = 0.37700, \hat{\theta}_2 = 0.24624, \hat{\sigma}_v = 0.55269\right)$ and then taking the square roots of the variance estimates $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_3^2$. These standard errors appear in Table 9.7, along with the forecast intervals calculated using $t_{(0.975, 93)} = 1.9858$. The forecast intervals are relatively wide, including the possibility of negative as well as positive growth. The point forecasts by themselves do not convey the great deal of uncertainty that is associated with these forecasts. Notice also how the forecast standard errors and the widths of the intervals increase as we forecast further into the future, reflecting the additional uncertainty from doing so.

9.7.2 Forecasting with an ARDL Model

In the previous section we saw how an autoregressive model can be used for forecasting, delivering both point and interval forecasts for a variable of interest. Suppose now that we wish to use an ARDL model for forecasting. As an example, consider forecasting future unemployment using the Okun's Law ARDL(1,1) model that we estimated in Section 9.6.2:

$$DU_t = \delta + \theta_1 DU_{t-1} + \delta_0 G_t + \delta_1 G_{t-1} + v_t \tag{9.69}$$

Does using this model for forecasting, instead of a pure AR model, create any special problems? One obvious difference is that future values of G are required. The value of DU in the first post-sample quarter is

$$DU_{T+1} = \delta + \theta_1 DU_T + \delta_0 G_{T+1} + \delta_1 G_T + v_{T+1} \tag{9.70}$$

Before we can use this equation to forecast DU_{T+1}, a value for G_{T+1} is needed; forecasting further into the future will require more future values of G. These values may be independent forecasts or they might be from "what if" questions: If GDP growth in the next two quarters is $G^*_{T+1} = G^*_{T+2}$, what is our forecast for the level of unemployment?

Apart from the need to supply future values of G, the forecasting procedure for an ARDL model is essentially the same as that for a pure AR model. Providing we are content to construct forecast intervals that ignore any error in the specification of future values of G, adding a distributed lag component to the AR model does not require any special treatment. Point and interval forecasts are obtained in the same way.

There is, however, one special feature of the model in (9.69) that is worthy of further consideration. Recall that the dependent variable DU_t is the *change* in unemployment defined as $DU_t = U_t - U_{t-1}$. Does this have any implications for forecasting the *level* of unemployment given by U_t? To investigate this question, we rewrite (9.70) as

$$U_{T+1} - U_T = \delta + \theta_1(U_T - U_{T-1}) + \delta_0 G_{T+1} + \delta_1 G_T + v_{T+1}$$

Bringing U_T over to the right side and collecting terms yields

$$\begin{aligned} U_{T+1} &= \delta + (\theta_1 + 1)U_T - \theta_1 U_{T-1} + \delta_0 G_{T+1} + \delta_1 G_T + v_{T+1} \\ &= \delta + \theta_1^* U_T + \theta_2^* U_{T-1} + \delta_0 G_{T+1} + \delta_1 G_T + v_{T+1} \end{aligned} \tag{9.71}$$

where $\theta_1^* = \theta_1 + 1$ and $\theta_2^* = -\theta_1$. For the purpose of computing point and interval forecasts, the ARDL(1,1) model for a *change* in unemployment can be written as an ARDL(2,1) model for the *level* of unemployment, with parameters θ_1^* and θ_2^*. This result holds not only for ARDL models where a dependent variable is measured in terms of a change or difference, but also for pure AR models involving such variables. It is particularly relevant when nonstationary variables are differenced to achieve stationarity—a transformation that is considered further in Chapter 12.

Finally, we note that forecasting with a finite distributed lag model with no AR component can be carried out within the same framework as forecasting (prediction) in the linear regression model considered in Chapter 6. Instead of the right-hand-side variables' being a number of different x's, they comprise a number of lags on the same x.

9.7.3 Exponential Smoothing

In Section 9.7.1 we saw how an autoregressive model can be used to forecast the future value of a variable by making use of past observations on that variable. Another popular model used for predicting the future value of a variable on the basis of its history is the exponential smoothing method. Like forecasting with an AR model, forecasting using exponential smoothing does not utilize information from any other variable.

To introduce this method, consider a sample of observations $(y_1, y_2, \ldots, y_{T-1}, y_T)$ where our objective is to forecast the next observation y_{T+1}. One possible forecasting method, and one that has some intuitive appeal, is to use the average of past information—say, the average of the last k observations. For example, if we adopt this method with $k = 3$, the proposed forecast is

$$\hat{y}_{T+1} = \frac{y_T + y_{T-1} + y_{T-2}}{3}$$

This forecasting rule is an example of a simple (equally-weighted) **moving average** model with $k = 3$. Note that when $k = 1$, all weight is placed on the most recent value and the forecast is $\hat{y}_{T+1} = y_T$.

Now let us extend the moving average idea by changing the equal weighting system where the weights are all $(1/k)$ to one where more weight is put on recent information—or, put another way, less weight is placed on observations further into the past. The exponential smoothing model is one such forecasting model; in this case, the weights decline exponentially as the observations get older. It has the form

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha)^1 y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \tag{9.72}$$

The weight attached to y_{T-s} is given by $\alpha(1-\alpha)^s$. We assume that $0 < \alpha \le 1$, which means that the weights get smaller as s gets larger (as we go further into the past). Also, using results on the infinite sum of a geometric progression, it can be shown that the weights sum to one: $\sum_{s=0}^{\infty}\alpha(1-\alpha)^s = 1$.

Using information from the infinite past is not convenient for forecasting. Recognizing that

$$(1-\alpha)\hat{y}_T = \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \alpha(1-\alpha)^3 y_{T-3} + \cdots. \tag{9.73}$$

allows us to simplify the model. Notice that the terms on the right hand side of (9.73) also appear on the right-hand side of (9.72). This means we can replace an infinite sum by a single term, so that the forecast can be more conveniently presented as

$$\hat{y}_{T+1} = \alpha y_T + (1-\alpha)\hat{y}_T \tag{9.74}$$

That is, the forecast for next period is a weighted average of the forecast for the current period and the actual realized value in the current period.

The exponential smoothing method is a versatile forecasting tool, but one needs a value for the smoothing parameter α and a value for $\hat{y}_T$ to generate the forecast $\hat{y}_{T+1}$. The value of α can reflect one's judgment about the relative weight of current information; alternatively, it can be estimated from historical information by obtaining **within-sample forecasts**

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1} \quad t = 2, 3, \ldots, T \tag{9.75}$$

and choosing that value of α which minimizes the sum of squares of the **one-step forecast errors**

$$v_t = y_t - \hat{y}_t = y_t - (\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}) \tag{9.76}$$

To compute $\sum_{t=2}^{T} v_t^2$ for a given value of α we need a starting value for $\hat{y}_1$. One option is to set $\hat{y}_1 = y_1$; another is to set $\hat{y}_1$ equal to the average of the first $(T+1)/2$ observations on y.[15] Once $\hat{y}_1$ has been set, (9.75) can be used recursively to generate a series of within-sample forecasts, and (9.76) can be used to generate a series of within-sample forecast errors.

The last value of the within-sample forecasts, generated, either with an α that reflects personal judgment or with one that minimizes $\sum_{t=2}^{T} v_t^2$, is $\hat{y}_T$, which is then used in (9.74) to generate a forecast for the first post-sample observation y_{T+1}. Forecasts for more than one period into the future are identical to that for period $T+1$. Can you see why?

For an illustration, we use the same quarterly data on U.S. GDP growth that was used for the AR model in Section 9.7.1. It runs from 1985Q2 to 2009Q3 and is stored in the file

[15] This second option is that used by the software EViews 7.0, and the one used in the example that follows.

okun.dat. Two values of α were chosen: $\alpha = 0.8$, and the value that minimized the sum of squares of within-sample forecast errors—in this case, $\alpha = 0.38$. The smaller the value of α, the greater the contribution of past observations to a forecast, and the smoother the series of within-sample forecasts is. With large values of α, the most recent observation is the major contributor to a forecast, and the series of forecasts more closely mimics the actual series. These characteristics are evident in Figures 9.12(a) and (b), where the actual series is graphed alongside the within-sample forecasts for $\alpha = 0.38$ (Figure 9.12(a)), and $\alpha = 0.8$ (Figure 9.12(b)). In both these figures, the solid line represents actual GDP growth and the dashed line represents the within-sample forecasts. In Figure 9.12(a), where $\alpha = 0.38$, the smoothed series is much less volatile than the actual series. It retains a jagged appearance, but the peaks and troughs are much less extreme. In Figure 9.12(b), where $\alpha = 0.8$, the peaks and troughs of the smoothed series are only slightly less pronounced, and the forecasts closely follow the actual series by one period reflecting the high weight placed on the most recent value.

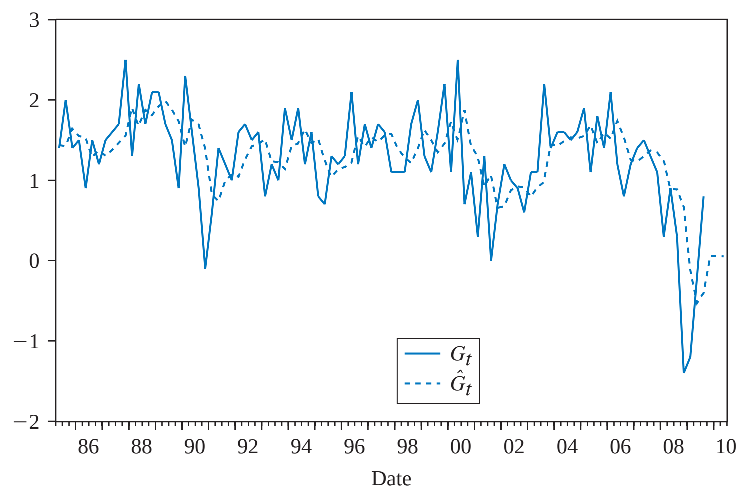

FIGURE 9.12 (a) Exponentially smoothed forecasts for GDP growth with $\alpha = 0.38$.

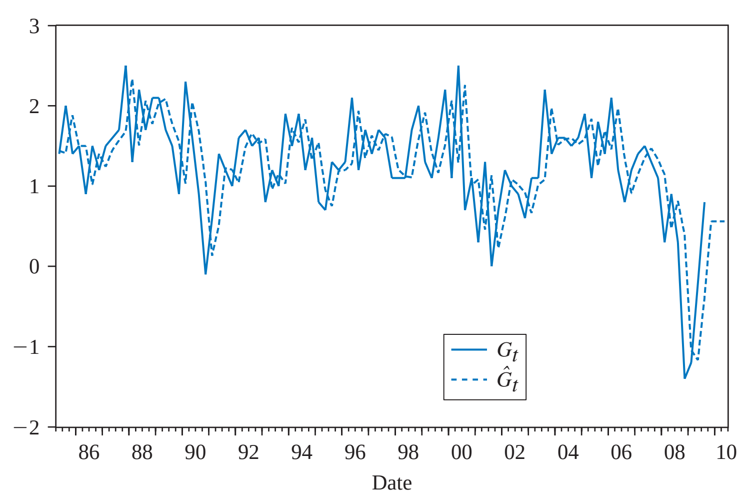

FIGURE 9.12 (b) Exponentially smoothed forecasts for GDP growth with $\alpha = 0.8$.

The forecasts for 2009Q4 from each value of α are

$$\alpha = 0.38: \quad \hat{G}_{T+1} = \alpha G_T + (1-\alpha)\hat{G}_T = 0.38 \times 0.8 + (1-0.38) \times (-0.403921)$$
$$= 0.0536$$

$$\alpha = 0.8: \quad \hat{G}_{T+1} = \alpha G_T + (1-\alpha)\hat{G}_T = 0.8 \times 0.8 + (1-0.8) \times (-0.393578)$$
$$= 0.5613$$

The difference between these two forecasts can be explained by the different weights placed on the most recent values of past growth. GDP growth was positive in 2009Q3 ($G_{2009Q3} = 0.8$) after three successive quarters of negative growth attributable to the global financial crisis ($G_{2009Q2} = -0.2$, $G_{2009Q1} = -1.2$, $G_{2008Q4} = -1.4$). The 2009Q4 forecast that uses $\alpha = 0.8$ is higher than that for $\alpha = 0.38$ because it places a heavy weight on the most recent positive growth in 2009Q3. The forecast for low growth that comes from using $\alpha = 0.38$ reflects the increased weight on the negative growth of the earlier three quarters.

9.8 Multiplier Analysis

Multiplier analysis refers to the effect, and the timing of the effect, of a change in one variable on the outcome of another variable. For example, by controlling the federal funds rate, the U.S. Federal Reserve Board attempts to influence inflation, unemployment, and the general level of economic activity. Because the effects of a change in the federal funds rate are not instantaneous, the Fed would like to know when and by how much variables like inflation and unemployment will respond. In a similar way, when the government makes changes to expenditure and taxation, it wants information on the magnitude and timing of changes in economic activity. At the firm level, firms are interested in the timing and magnitude of the effects of various forms of advertising on sales of their products.

The concepts of impact, delay, interim, and total multipliers were introduced in Section 9.2 in the context of a finite distributed lag model. If your memory needs refreshing, please reread that section. Now we are concerned with how to find multipliers for an ARDL model of the form

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + v_t \tag{9.77}$$

The secret for doing so lies in our ability to transform it into an infinite distributed lag model written as

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots + e_t \tag{9.78}$$

The multipliers defined for this model are similar to those for the finite distributed lag model. Specifically,

$$\beta_s = \frac{\partial y_t}{\partial x_{t-s}} = s \text{ period delay multiplier}$$

$$\sum_{j=0}^{s} \beta_j = s \text{ period interim multiplier}$$

$$\sum_{j=0}^{\infty} \beta_j = \text{total multiplier}$$

When we estimate an ARDL model, we obtain estimates of the θ's and δ's in (9.77). To obtain multipliers that are expressed in terms of the β's, we need to be able to compute estimates of the β's from those for the θ's and δ's. Describing how the β's can be derived from the θ's and δ's is the purpose of this section.

Our task is made easier if we can master some machinery known as the **lag operator**. The lag operator L has the effect of lagging a variable,

$$Ly_t = y_{t-1}$$

For lagging a variable twice, we have

$$L(Ly_t) = Ly_{t-1} = y_{t-2}$$

which we write as $L^2 y_t = y_{t-2}$. More generally, L raised to the power of s means lag a variable s times

$$L^s y_t = y_{t-s}$$

Now we are in a position to write the ARDL model in terms of lag operator notation. Equation (9.77) becomes

$$\begin{aligned} y_t = \delta + \theta_1 L y_t + \theta_2 L^2 y_t + \cdots + \theta_p L^p y_t + \delta_0 x_t + \delta_1 L x_t + \delta_2 L^2 x_t \\ + \cdots + \delta_q L^q x_t + v_t \end{aligned} \tag{9.79}$$

Bringing the terms that contain y_t to the left side of the equation, and factoring out y_t and x_t yields

$$\left(1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p\right) y_t = \delta + \left(\delta_0 + \delta_1 L + \delta_2 L^2 + \cdots + \delta_q L^q\right) x_t + v_t \tag{9.80}$$

This algebra is starting to get heavy. To make our derivation manageable, consider the ARDL(1,1) model used to describe Okun's law. From the above results, the model

$$DU_t = \delta + \theta_1 DU_{t-1} + \delta_0 G_t + \delta_1 G_{t-1} + v_t \tag{9.81}$$

can be written as

$$(1 - \theta_1 L) DU_t = \delta + (\delta_0 + \delta_1 L) G_t + v_t \tag{9.82}$$

Now suppose that it is possible to define an inverse of $(1 - \theta_1 L)$, which we write as $(1 - \theta_1 L)^{-1}$, which is such that

$$(1 - \theta_1 L)^{-1} (1 - \theta_1 L) = 1$$

This concept is a bit abstract. Using it will seem like magic the first time that you encounter it. Stick with us. We have nearly reached the essential result. Multiplying both sides of (9.82) by $(1 - \theta_1 L)^{-1}$ yields

$$DU_t = (1 - \theta_1 L)^{-1} \delta + (1 - \theta_1 L)^{-1} (\delta_0 + \delta_1 L) G_t + (1 - \theta_1 L)^{-1} v_t \tag{9.83}$$

This representation is useful because we can equate it with the infinite distributed lag representation

$$\begin{aligned} DU_t &= \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \beta_3 G_{t-3} + \cdots + e_t \\ &= \alpha + \left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\right) G_t + e_t \end{aligned} \tag{9.84}$$

For (9.83) and (9.84) to be identical, it must be true that

$$\alpha = (1 - \theta_1 L)^{-1}\delta \tag{9.85}$$

$$\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots = (1 - \theta_1 L)^{-1}(\delta_0 + \delta_1 L) \tag{9.86}$$

$$e_t = (1 - \theta_1 L)^{-1} v_t \tag{9.87}$$

Equation (9.85) can be used to derive α in terms of θ_1 and δ, and (9.86) can be used to derive the β's in terms of the θ's and δ's. To see how, first multiply both sides of (9.85) by $(1 - \theta_1 L)$ to obtain $(1 - \theta_1 L)\alpha = \delta$. Then, recognizing that the lag of a constant that does not change over time is the same constant $(L\alpha = \alpha)$, we have

$$(1 - \theta_1)\alpha = \delta \quad \text{and} \quad \alpha = \frac{\delta}{1 - \theta_1}$$

Turning now to the β's, we multiply both sides of (9.86) by $(1 - \theta_1 L)$ to obtain

$$\begin{aligned} \delta_0 + \delta_1 L &= (1 - \theta_1 L)\left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\right) \\ &= \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots \\ &\qquad - \beta_0\theta_1 L - \beta_1\theta_1 L^2 - \beta_2\theta_1 L^3 - \cdots \\ &= \beta_0 + (\beta_1 - \beta_0\theta_1)L + (\beta_2 - \beta_1\theta_1)L^2 + (\beta_3 - \beta_2\theta_1)L^3 + \cdots \end{aligned} \tag{9.88}$$

Notice how we can do algebra with the lag operator. We have used the fact that $L^r L^s = L^{r+s}$.

Equation (9.88) holds the key to deriving the β's in terms of the θ's and the δ's. For both sides of this equation to mean the same thing (to imply the same lags), coefficients of like powers in the lag operator must be equal. To make what follows more transparent, we rewrite (9.88) as

$$\delta_0 + \delta_1 L + 0L^2 + 0L^3 = \beta_0 + (\beta_1 - \beta_0\theta_1)L + (\beta_2 - \beta_1\theta_1)L^2 + (\beta_3 - \beta_2\theta_1)L^3 + \cdots \tag{9.89}$$

Equating coefficients of like powers in L yields

$$\begin{aligned} \delta_0 &= \beta_0 \\ \delta_1 &= \beta_1 - \beta_0\theta_1 \\ 0 &= \beta_2 - \beta_1\theta_1 \\ 0 &= \beta_3 - \beta_2\theta_1 \end{aligned}$$

and so on. Thus, the β's can be found from the θ's and the δ's using the recursive equations

$$\begin{aligned} \beta_0 &= \delta_0 \\ \beta_1 &= \delta_1 + \beta_0\theta_1 \\ \beta_j &= \beta_{j-1}\theta_1 \ \text{ for } \ j \geq 2 \end{aligned} \tag{9.90}$$

You are probably asking: Do I have to go through all this each time I want to derive some multipliers for an ARDL model? The answer is no. You can start from the equivalent of (9.88) which, in its general form, is

$$\delta_0 + \delta_1 L + \delta_2 L^2 + \cdots + \delta_q L^q = \left(1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p\right) \times \left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\right) \tag{9.91}$$

Given the values p and q for your ARDL model, you need to multiply out the above expression, and then equate coefficients of like powers in the lag operator.

What are the values of the multipliers for our Okun's Law example?

$$\widehat{DU}_t = 0.3780 + 0.3501 DU_{t-1} - 0.1841 G_t - 0.0992 G_{t-1}$$

Using the relationships in (9.90), the impact multiplier and the delay multipliers for the first four quarters are given by

$$\hat{\beta}_0 = \hat{\delta}_0 = -0.1841$$

$$\hat{\beta}_1 = \hat{\delta}_1 + \hat{\beta}_0\hat{\theta}_1 = -0.099155 - 0.184084 \times 0.350116 = -0.1636$$

$$\hat{\beta}_2 = \hat{\beta}_1\hat{\theta}_1 = -0.163606 \times 0.350166 = -0.0573$$

$$\hat{\beta}_3 = \hat{\beta}_2\hat{\theta}_1 = -0.057281 \times 0.350166 = -0.0201$$

$$\hat{\beta}_4 = \hat{\beta}_3\hat{\theta}_1 = -0.020055 \times 0.350166 = -0.0070$$

An increase in GDP growth leads to a fall in unemployment, with its greatest effect being felt in the current and next quarters and a declining effect thereafter. The effect eventually declines to zero. This property—that the weights at long lags go to zero—is an essential one for the above analysis to be valid. The weights are displayed in Figure 9.13 for lags up to seven quarters.

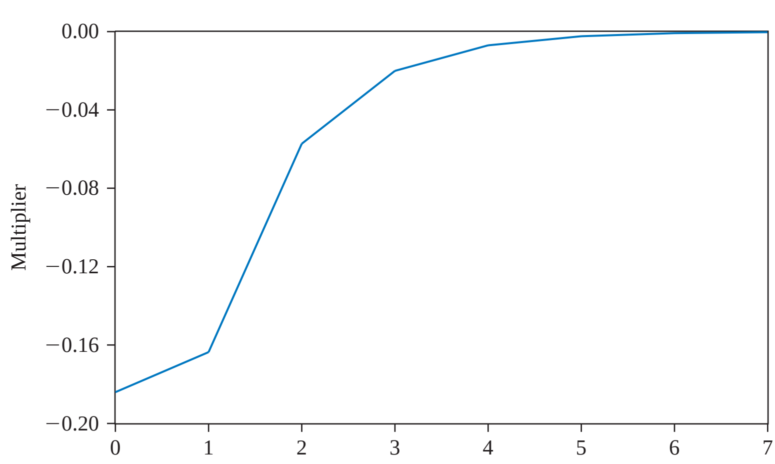

FIGURE 9.13 Delay multipliers from Okun's law ARDL(1,1) model.

Finally, we can estimate the total multiplier that is given by $\sum_{j=0}^{\infty}\beta_j$, and the normal growth rate that is needed to maintain a constant rate of unemployment, $G_N = -\alpha\big/\sum_{j=0}^{\infty}\beta_j$. The total multiplier can be found by summing those β's which are sufficiently large to contribute to the sum or by using results on the sum of an infinite geometric progression. For the latter approach, we can show that

$$\sum_{j=0}^{\infty}\hat{\beta}_j = \hat{\delta}_0 + \frac{\hat{\delta}_1 + \hat{\delta}_0\hat{\theta}_1}{1-\hat{\theta}_1} = -0.184084 + \frac{-0.163606}{1-0.350116} = -0.4358$$

An estimate for α is given by $\hat{\alpha} = \hat{\delta}\big/\left(1-\hat{\theta}_1\right) = 0.37801/0.649884 = 0.5817$ which leads to a normal growth rate of $\hat{G}_N = 0.5817/0.4358 = 1.3\%$ per quarter. These results are consistent with those that we found from the finite distributed lag model in Section 9.2. In that instance, we had -0.437 for the total multiplier and 1.3% for G_N.

9.9 Exercises

Answers to exercises marked * appear at www.wiley.com/college/hill.

9.9.1 Problems

9.1 Consider the following distributed lag model relating the percentage growth in private investment (*INVGWTH*) to the federal funds rate of interest (*FFRATE*):

$$\begin{aligned} INVGWTH_t = 4 &- 0.4FFRATE_t - 0.8FFRATE_{t-1} - 0.6FFRATE_{t-2} \\ &- 0.2FFRATE_{t-3} \end{aligned}$$

(a) Suppose *FFRATE* = 1% for $t = 1, 2, 3, 4$. Use the above equation to forecast *INVGWTH* for $t = 4$.

(b) Suppose *FFRATE* is raised to 1.5% in period $t = 5$ and then returned to its original level of 1% for $t = 6, 7, 8, 9$. Use the equation to forecast *INVGWTH* for periods $t = 5, 6, 7, 8, 9$. Relate the changes in your forecasts to the values of the coefficients. What are the delay multipliers?

(c) Suppose *FFRATE* is raised to 1.5% for periods $t = 5, 6, 7, 8, 9$. Use the equation to forecast *INVGWTH* for periods $t = 5, 6, 7, 8, 9$. Relate the changes in your forecasts to the values of the coefficients. What are the interim multipliers? What is the total multiplier?

9.2 The file *ex9_2.dat* contains 105 weekly observations on sales revenue (*SALES*) and advertising expenditure (*ADV*) in millions of dollars for a large midwest department store in 2008 and 2009. The following relationship was estimated:

$$\widehat{SALES_t} = 25.34 + 1.842\ ADV_t + 3.802\ ADV_{t-1} + 2.265\ ADV_{t-2}$$

(a) Describe the relationship between sales and advertising expenditure. Include an explanation of the lagged relationship. When does advertising have its greatest impact? What is the total effect of a sustained \$1 million increase in advertising expenditure?

(b) The estimated covariance matrix of the coefficients is

	C	ADV	ADV_{t-1}	ADV_{t-2}
C	2.5598	−0.7099	−0.1317	−0.7661
ADV	−0.7099	1.3946	−1.0406	0.0984
ADV_{t-1}	−0.1317	−1.0406	2.1606	−1.0367
ADV_{t-2}	−0.7661	0.0984	−1.0367	1.4214

Using a one-tail test and a 5% significance level, which lag coefficients are significantly different from zero? Do your conclusions change if you use a one-tail test? Do they change if you use a 10% significance level?

(c) Find 95% confidence intervals for the impact multiplier, the one-period interim multiplier, and the total multiplier.

9.3 Reconsider the estimated equation and covariance matrix in Exercise 9.2. Suppose, as a marketing executive for the department store, that you have a total of \$6 million to spend on advertising over the next three weeks, $t = 106$, 107, and 108. Consider the following allocations of the \$6 million:

$$ADV_{106} = 6, \quad ADV_{107} = 0, \quad ADV_{108} = 0$$
$$ADV_{106} = 0, \quad ADV_{107} = 6, \quad ADV_{108} = 0$$
$$ADV_{106} = 2, \quad ADV_{107} = 4, \quad ADV_{108} = 0$$

(a) For each allocation of the \$6 million, forecast sales revenue for $t = 106$, 107, and 108. Which allocation leads to the largest forecast for total sales revenue over the three weeks? Which allocation leads to the largest forecast for sales in week $t = 108$? Explain why these outcomes were obtained.

(b) Find 95% forecast intervals for ADV_{108} for each of the three allocations. If maximizing ADV_{108} is your objective, which allocation would you choose? Why?

9.4* The following least squares residuals come from a sample of size $T = 10$:

t	1	2	3	4	5	6	7	8	9	10
$\hat{e}_t$	0.28	−0.31	−0.09	0.03	−0.37	−0.17	−0.39	−0.03	0.03	1.02

(a) Use a hand calculator to compute the sample autocorrelations:

$$r_1 = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} \qquad r_2 = \frac{\sum_{t=3}^{T} \hat{e}_t \hat{e}_{t-2}}{\sum_{t=1}^{T} \hat{e}_t^2}$$

(b) Test whether (i) r_1 is significantly different from zero and (ii) r_2 is significantly different from zero. Sketch the first two bars of the correlogram. Include the significance bounds.

9.5 The file *growth47.dat* contains 250 quarterly observations on U.S. GDP growth from quarter two, 1947, to quarter three, 2009. From these data, we calculate the following quantities:

$$\sum_{t=1}^{250}\left(G_t-\overline{G}\right)^2=333.8558 \qquad \sum_{t=2}^{250}\left(G_t-\overline{G}\right)\left(G_{t-1}-\overline{G}\right)=162.9753$$

$$\sum_{t=3}^{250}\left(G_t-\overline{G}\right)\left(G_{t-2}-\overline{G}\right)=112.4882 \qquad \sum_{t=4}^{250}\left(G_t-\overline{G}\right)\left(G_{t-3}-\overline{G}\right)=30.5802$$

(a) Compute the first three autocorrelations (r_1, r_2 and r_3) for G. Test whether each one is significantly different from zero at a 5% significance level. Sketch the first three bars of the correlogram. Include the significance bounds.

(b) Given that $\sum_{t=2}^{250}\left(G_{t-1}-\overline{G}_{-1}\right)^2=333.1119$ and $\sum_{t=2}^{250}\left(G_t-\overline{G}_1\right)\left(G_{t-1}-\overline{G}_{-1}\right)=162.974$, where $\overline{G}_1=\sum_{t=2}^{250}G_t/249=1.662249$ and $\overline{G}_{-1}=\sum_{t=2}^{250}G_{t-1}/249=1.664257$, find least squares estimates of δ_1 and θ_1 in the AR(1) model $G_t=\delta+\theta_1 G_{t-1}+e_t$. Explain the difference between the estimate $\hat{\theta}_1$ and the estimate r_1 obtained in part (a).

9.6 Increases in the mortgage interest rate increase the cost of owning a house and lower the demand for houses. In this question we consider an equation where the monthly change in the number of new one-family houses sold in the U.S. depends on last month's change in the 30-year conventional mortgage rate. Let *HOMES* be the number of new houses sold (in thousands) and *IRATE* be the mortgage rate. Their monthly changes are denoted by $DHOMES_t = HOMES_t - HOMES_{t-1}$ and $DIRATE_t = IRATE_t - IRATE_{t-1}$. Using data from January 1992 to March 2010 (stored in the file *homes.dat*), we obtain the following least squares regression estimates:

$$\widehat{DHOMES_t} = \underset{(3.498)}{-2.077} - \underset{(16.98)}{53.51}DIRATE_{t-1} \qquad \text{obs} = 218$$
$$\text{(se)}$$

(a) Interpret the estimate -53.51. Construct and interpret a 95% confidcnce interval for the coefficient of $DIRATE_{t-1}$.

(b) Let $\hat{e}_t$ denote the residuals from the above equation. Use the following estimated equation to conduct two separate tests for first-order autoregressive errors.

$$\hat{e}_t = -0.1835 - \underset{(16.087)}{3.210}DIRATE_{t-1} - \underset{(0.0649)}{0.3306}\hat{e}_{t-1} \qquad R^2 = 0.1077$$
$$\text{(se)} \qquad \text{obs} = 218$$

(c) The model with AR(1) errors was estimated as

$$\widehat{DHOMES_t} = \underset{(2.497)}{-2.124} - \underset{(14.10)}{58.61}DIRATE_{t-1} \qquad e_t = \underset{(0.0649)}{-0.3314}e_{t-1} + \hat{v}_t$$
$$\text{(se)} \qquad \text{obs} = 217$$

Construct a 95% confidence interval for the coefficient of $DIRATE_{t-1}$, and comment on the effect of ignoring autocorrelation on inferences about this coefficient.

9.7* Consider the model

$$e_t = \rho e_{t-1} + v_t$$

(a) Suppose $\rho = 0.9$ and $\sigma_v^2 = 1$. What is (i) the correlation between e_t and e_{t-1}? (ii) the correlation between e_t and e_{t-4}? (iii) the variance σ_e^2?

(b) Repeat part (a) with $\rho = 0.4$ and $\sigma_v^2 = 1$. Comment on the difference between your answers for parts (a) and (b).

9.8 In Section 9.1, the following Phillips curve was estimated:

$$\widehat{INF}_t = 0.1001 + 0.2354\, INF_{t-1} + 0.1213\, INF_{t-2} + 0.1677\, INF_{t-3} + 0.2819\, INF_{t-4} - 0.7902\, DU_t$$

The last four sample values for inflation are $INF_{2009Q3} = 1.0$, $INF_{2009Q2} = 0.5$, $INF_{2009Q1} = 0.1$, and $INF_{2008Q4} = -0.3$. The unemployment rate in 2009Q3 was 5.8%. The estimated error variance for the above equation is $\hat{\sigma}_v^2 = 0.225103$.

(a) Given that the unemployment rates in the first three post-sample quarters are $U_{2009Q4} = 5.6$, $U_{2010Q1} = 5.4$, and $U_{2010Q2} = 5.0$, use the estimated equation to forecast inflation for 2009Q4, 2010Q1 and 2010Q2.

(b) Find the standard errors of the forecast errors for your forecasts in (a).

(c) Find 95% forecast intervals for $INF_{2009Q4}, INF_{2010Q1}$, and INF_{2010Q2}. How reliable are the forecasts you found in part (a)?

9.9 Consider the infinite lag representation

$$y_t = \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + e_t$$

for the ARDL model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \theta_4 y_{t-4} + \delta_0 x_t + v_t$$

(a) Show that

$$\begin{aligned}
\alpha &= \delta/(1 - \theta_1 - \theta_2 - \theta_3 - \theta_4) \\
\beta_0 &= \delta_0 \\
\beta_1 &= \beta_0\theta_0 \\
\beta_2 &= \beta_1\theta_1 + \beta_0\theta_2 \\
\beta_3 &= \beta_2\theta_1 + \beta_1\theta_2 + \beta_0\theta_3 \\
\beta_s &= \beta_{s-1}\theta_1 + \beta_{s-2}\theta_2 + \beta_{s-3}\theta_3 + \beta_{s-4}\theta_4 \quad \text{for } s \geq 4
\end{aligned}$$

(b) Use the results in (a) to find estimates of the first 12 lag weights for the estimated Phillips curve in Exercise 9.8. Graph those weights and comment on the graph.

(c) What rate of inflation is consistent with a constant unemployment rate (where $DU = 0$ in all time periods)?

9.10* Quarterly data from 1960Q1 to 2009Q4, stored in the file *consumptn.dat*, were used to estimate the following relationship between growth in consumption of consumer durables in the U.S. (*DURGWTH*) and growth in personal disposable income (*INCGWTH*):

$$\widehat{DURGWTH_t} = 0.0103 - 0.1631DURGWTH_{t-1} + 0.7422INCGWTH_t \\ + 0.3479INCGWTH_{t-1}$$

(a) Given that $DURGWTH_{2009Q4} = 0.1$, $INCGWTH_{2009Q4} = 0.9$, $INCGWTH_{2010Q1} = 0.6$, and $INCGWTH_{2010Q2} = 0.8$, forecast *DURGWTH* for 2010Q1 and 2010Q2.
(b) Find and comment on the implied lag weights for up to 12 quarters for the infinite distributed lag representation

$$DURGWTH_t = \alpha + \sum_{s=0}^{\infty} \beta_s INCGWTH_{t-s} + e_t$$

(c) Find values for the one- and two-quarter delay and interim multipliers, and the total multiplier. Interpret those values.

9.11 (a) Write the AR(1) error model $e_t = \rho e_{t-1} + v_t$ in lag operator notation.
(b) Show that

$$(1 - \rho L)^{-1} = 1 + \rho L + \rho^2 L + \rho^3 L^3 + \cdots$$

and hence that

$$e_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots$$

9.9.2 Computer Exercises

9.12* Consider the Okun's Law finite distributed lag model that was estimated in Section 9.2 and the data for which appears in *okun.dat*.
(a) Estimate the following model for q = 0, 1, 2, 3, 4, 5, and 6.

$$DU_t = \alpha + \sum_{s=0}^{q} \beta_s G_{t-s} + e_t$$

In each case use data from t = 1986Q4 to t = 2009Q3 to ensure that 92 observations are used for each estimation. Report the values of the AIC and SC selection criteria for each value of q. What lag length would you choose on the basis of the AIC? What lag length would you choose based on the SC?
(b) Using the model that minimizes the AIC:

(i) Find a 95% confidence interval for the impact multiplier.
(ii) Test the null hypothesis that the total multiplier equals –0.5 against the alternative that it is greater than –0.5. Use a 5% significance level.
(iii) Find a 95% confidence interval for the normal growth rate G_N. (*Hint*: Use your software to get the standard error for $\hat{G}_N = \hat{\alpha}/\hat{\gamma}$ where $\hat{\gamma} = -\Sigma_{s=0}^{q} b_s$. You can do so by pretending to test a hypothesis such as $H_0 : \alpha/\gamma = 1$.)

9.13 The file *ex9_13.dat* contains 157 weekly observations on sales revenue (*SALES*) and advertising expenditure (*ADV*) in millions of dollars for a large midwest department store for 2005–2007. (Exercise 9.2 used data on this store for 2008–2009.) The weeks are from December 28, 2004, to December 25, 2007. We denote them as $t = 1, 2, \ldots, 157$.

(a) Graph the series for *SALES* and *ADV*. Do they appear be trending or do they appear to fluctuate around a constant mean? On your graphs, draw horizontal lines at the means of the series.

(b) Estimate a finite distributed lag model of the form

$$SALES_t = \alpha + \sum_{s=0}^{q} \beta_s ADV_{t-s} + e_t$$

for $q = 0, 1, 2, 3, 4$, and 5. In each case use 152 observations $(t = 6, 7, \ldots, 157$ where $t = 6$ is February 1, 2005). Report the SC values and the total multipliers for each equation. Is the estimated total multiplier sensitive to choice of lag length?

(c) Comment on the estimated lag structure of the model that minimizes the SC. Does it seem sensible to you? Are all the estimates significantly different from zero at a 5% significance level? Use this model to answer the remaining parts of this question.

(d) Construct 95% interval estimates for the (i) one-week delay multiplier, (ii) one-week interim multiplier, (iii) two-week delay multiplier, and (iv) two-week interim multiplier.

(e) The CEO claims that increasing advertising expenditure by \$1 million a week in each of the next three weeks will increase total sales over those three weeks by at least \$6 million. Is there enough evidence in the data to support this claim?

(f) Forecast sales revenue for the first four post-sample weeks, $t = 158, 159, 160, 161$ when (i) nothing is spent on advertising for those four weeks, (ii) \$4 million is spent in the first week ($t = 158$), and nothing is spent in the remaining three weeks, and (iii) \$1 million is spent in each of the four weeks. Comment on the three different forecast paths.

9.14 One way of modeling supply response for an agricultural crop is to specify a model in which area planted (acres) depends on price. When the price of the crop's output is high, farmers plant more of that crop than when its price is low. Letting *AREA* denote area planted, and *PRICE* denote output price, and assuming a log-log (constant elasticity) functional form, a finite distributed lag area response model of this type can be written as

$$\ln(AREA_t) = \alpha + \sum_{s=0}^{q} \beta_s \ln(PRICE_{t-s}) + e_t$$

We use this model to explain the area of sugar cane planted in a region of the southeast Asian country of Bangladesh. Information on the delay and interim elasticities is useful for government planning. It is important to know whether existing sugar processing mills are likely to be able to handle predicted output, whether there is likely to be excess milling capacity, and whether a pricing policy linking production, processing, and consumption is desirable. Data comprising 34 annual observations on area and price are given in the file *bangla.dat*.

(a) Estimate this model assuming $q = 4$. What are the estimated delay and interim elasticities? Comment on the results.

(b) You will have discovered that the lag weights obtained in part (a) are not sensible. One way to try and overcome this problem is to insist that the weights lie on a straight line

$$\beta_s = \alpha_0 + \alpha_1 s \qquad s = 0,\ 1,\ 2,\ 3,\ 4$$

If $\alpha_0 > 0$ and $\alpha_1 < 0$, these weights will decline, implying that farmers place a larger weight on more recent prices when forming their expectations. Substitute $\beta_s = \alpha_0 + \alpha_1 s$ into the original equation and hence show that this equation can be written as

$$\ln(AREA_t) = \alpha + \alpha_0 z_{t0} + \alpha_1 z_{t1} + e_t$$

where $z_{t0} = \sum_{s=0}^{4} \ln(PRICE_{t-s})$ and $z_{t1} = \sum_{s=1}^{4} s \ln(PRICE_{t-s})$.

(c) Create variables z_{t0} and z_{t1} and find least squares estimates of α_0 and α_1.
(d) Use the estimates for α_0 and α_1 to find estimates for $\beta_s = \alpha_0 + \alpha_1 s$ and comment on them. Has the original problem been cured? Do the weights now satisfy a priori expectations?
(e) How do the delay and interim elasticities compare with those obtained earlier?

9.15* Reconsider the sugar cane supply response problem that was introduced in Exercise 9.14. Using data in *bangla.dat*, estimate the following model with no lags

$$\ln(AREA_t) = \beta_1 + \beta_2 \ln(PRICE_t) + e_t$$

(a) Find the correlogram for the residuals. What autocorrelations are significantly different from zero?
(b) Perform an *LM* test for autocorrelated errors using one lagged residual and a 5% significance level.
(c) Find two 95% confidence intervals for the elasticity of supply—one using least squares standard errors and one using HAC standard errors. What are the consequences for interval estimation when serially correlated errors are ignored?
(d) Estimate the model under the assumption that the error is an AR(1) process. Is the estimate for ρ significantly different from zero at a 5% significance level? Compute a 95% confidence interval for the elasticity of supply. How does it compare with those obtained in part (c)?
(e) Estimate an ARDL(1,1) model for sugar supply response. What restrictions are necessary on the coefficients of this model to make it equivalent to that in (d)? Test these restrictions using a 5% significance level. Do the residuals from this model show any evidence of serial correlation?

9.16* Consider further the ARDL(1,1) supply response model for sugar cane estimated in part (e) of Exercise 9.15.

$$\ln(AREA_t) = \delta + \theta_1 \ln(AREA_{t-1}) + \delta_0 \ln(PRICE_t) + \delta_1 \ln(PRICE_{t-1}) + v_t$$

(a) Suppose the first two post-sample prices are $PRICE_{T+1} = 1$ and $PRICE_{T+2} = 0.8$. Use the estimated equation to forecast $\ln(AREA)$ in years $T + 1$ and $T + 2$. What are the corresponding forecasts for *AREA*?
(b) Find 95% forecast intervals for *AREA* in years $T + 1$ and $T + 2$. Can we forecast area accurately?
(c) Use the results in (9.90) and the estimated equation to find lag and interim elasticities for up to four years. Interpret these values.
(d) Find the estimated total elasticity. What does this value tell you?

9.17 The file *growth47.dat* contains 250 quarterly observations on U.S. GDP growth (percentage change in GDP) from quarter 2, 1947, to quarter 3, 2009.

(a) Estimate an AR(2) model for GDP growth and check to see if the residuals are autocorrelated. What residual autocorrelations, if any, are significantly different from zero? Does an *LM* test with two lagged errors suggest serially correlated errors?
(b) Repeat part(a) using an AR(3) model.
(c) Use the estimated AR(3) model to find 95% forecast intervals for growth in 2009Q4, 2010Q1, and 2010Q2. Check to see if the actual growth figures fell within your forecast intervals. (You can find these figures on the Federal Reserve Economic Data (FRED) web page maintained by the Federal Reserve Bank of St. Louis).

9.18 You wish to compare the performance of an AR model and an exponential smoothing model for forecasting sales revenue one week into the future.
(a) Using the data in *ex9_13.dat*, estimate an AR(2) model for *SALES*. Check to see if the errors are serially correlated.
(b) Re-estimate the AR(2) model with the last four observations ($t = 154$, 155, 156, and 157) omitted. Use the estimated model to forecast *SALES* for $t = 154$ (one week ahead). Call the forecast $SALES_{154}^{AR}$.
(c) Re-estimate the AR(2) model with the last three observations ($t = 155$, 156, and 157) omitted. Use the estimated model to forecast *SALES* for $t = 155$ (one week ahead). Call the forecast $SALES_{155}^{AR}$.
(d) Continue the process described in parts (b) and (c) to obtain forecasts $SALES_{156}^{AR}$ and $SALES_{157}^{AR}$.
(e) Follow the same procedure with an exponential smoothing model. First with the last four observations omitted, then the last three, then the last two, and then the last one, find the smoothing parameter estimate which minimizes the sum of squares of the within-sample one-step forecast errors. In each case use the estimated smoothing parameter to forecast one week ahead, obtaining the forecasts $SALES_{154}^{ES}$, $SALES_{155}^{ES}$, $SALES_{156}^{ES}$ and $SALES_{157}^{ES}$.
(f) Find the mean-square prediction errors (MSPE) $\sum_{t=154}^{157}\left(SALES_t^{AR} - SALES_t\right)^2/4$ and $\sum_{t=154}^{157}\left(SALES_t^{ES} - SALES_t\right)^2/4$. On the basis of their MSPEs, which method has led to the most accurate forecasts?

9.19 In this exercise we explore further the relationship between houses sold and the mortgage rate that was introduced in Exercise 9.6. To familiarize yourself with the variables, go back and read the question for Exercise 9.6. Then, use the data in *homes.dat* to answer the following questions:
(a) Graph *HOMES*, *IRATE*, *DHOMES*, and *DIRATE*. Which variables appear to be trending? Which ones are not trending?
(b) Estimate the following model and report the results. Are all the estimates significantly different from zero at a 5% significance level?

$$\begin{aligned} DHOMES_t = \delta + \theta_1 DHOMES_{t-1} + \delta_1 DIRATE_{t-1} \\ + \delta_2 DIRATE_{t-2} + v_t \end{aligned} \tag{9.92}$$

(c) Test the hypothesis $H_0: \theta_1\delta_1 = -\delta_2$ against the alternative $H_1: \theta_1\delta_1 \neq -\delta_2$ at a 5% significance level. What does the outcome of this test tell you?
(d) Find the correlogram of the residuals from estimating (9.92). Does it show any evidence of serial correlation in the errors?

(e) Test for serially correlated errors in (9.92) using an *LM* test with two lagged errors.

(f) Estimate the following ARDL model:

$$DHOMES_t = \delta + \theta_1 DHOMES_{t-1} + \theta_5 DHOMES_{t-5} + \delta_1 DIRATE_{t-1} + \delta_3 DIRATE_{t-3} + v_t \tag{9.93}$$

This is a special case of an ARDL(5,3) model where $\theta_2 = \theta_3 = \theta_4 = \delta_0 = \delta_2 = 0$. Is this equation an improvement over (9.92)? Why?

9.20 (a) Show that (9.93) can be written as

$$HOMES_t = \delta + (\theta_1 + 1)HOMES_{t-1} - \theta_1 HOMES_{t-2} + \theta_5 DHOMES_{t-5} + \delta_1 DIRATE_{t-1} + \delta_3 DIRATE_{t-3} + v_t$$

(b) If you have not already done so, estimate (9.93). Use this estimated equation and the result in part (a) to forecast the number of new one-family houses sold in April, May, and June 2010, assuming the mortgage rate in those three months remains constant at 4.97%

(c) Find 95% forecast intervals for the three forecasts made in part (b).

9.21 In (9.59) we obtained the following estimated equation for Okun's Law

$$\widehat{DU}_t = \underset{(0.0578)}{0.3780} + \underset{(0.0846)}{0.3501}DU_{t-1} - \underset{(0.0307)}{0.1841}G_t - \underset{(0.0368)}{0.0992}G_{t-1}$$
$$\text{(se)}$$

(a) Use the data in *okun.dat* to reproduce these estimates.

(b) Check the correlogram of the residuals. Are there any significant autocorrelations?

(c) Carry out *LM* tests for autocorrelation on the residuals for error lags up to four.

(d) Re-estimate the equation with variables DU_{t-2} and G_{t-2} added separately and then together. Are their coefficients significantly different from zero?

(e) What do you conclude about the specification in (9.59)?

9.22 An important relationship in macroeconomics is the consumption function. The file *consumptn.dat* contains quarterly data from 1960Q1 to 2009Q4 on the percentage changes in disposable personal income and personal consumption expenditures. We describe these variables as income growth (*INCGWTH*) and consumption growth (*CONGWTH*). To ensure that the same number of observations (197) are used for estimation in each of the models that we consider, use as your sample period 1960Q4 to 2009Q4. Where relevant, lagged variables on the right-hand side of equations can use values prior to 1960Q4.

(a) Graph the time series for *CONGWTH* and *INCGWTH*. Include a horizontal line at the mean of each series. Do the series appear to fluctuate around a constant mean?

(b) Estimate the model $CONGWTH_t = \delta + \delta_0 INCGWTH_t + v_t$. Interpret the estimate for δ_0. Check for serially correlated errors using the residual correlogram, and an *LM* test with two lagged errors. What do you conclude?

(c) Estimate the model $CONGWTH_t = \delta + \theta_1 CONGWTH_{t-1} + \delta_0 INCGWTH_t + v_t$. Is this model an improvement over that in part (b)? Is the estimate for θ_1

significantly different from zero? Have the values for the AIC and the SC gone down? Has serial correlation in the errors been eliminated?

(d) Add the variable $CONGWTH_{t-2}$ to the model in part (c) and re-estimate. Is this model an improvement over that in part (c)? Is the estimate for θ_2 (the coefficient of $CONGWTH_{t-2}$) significantly different from zero? Have the values for the AIC and the SC gone down? Has serial correlation in the errors been eliminated?

(e) Add the variable $INCGWTH_{t-1}$ to the model in part (d) and re-estimate. Is this model an improvement over that in part (d)? Is the estimate for δ_1 (the coefficient of $INCGWTH_{t-1}$) significantly different from zero? Have the values for the AIC and the SC gone down? Has serial correlation in the errors been eliminated?

(f) Does the addition of $CONGWTH_{t-3}$ or $INCGWTH_{t-2}$ improve the model in part (e)?

(g) Drop the variable $CONGWTH_{t-1}$ from the model in part (e) and re-estimate. Why might you consider dropping this variable? The model you should be estimating is

$$\begin{aligned} CONGWTH_t = \delta + \theta_2 CONGWTH_{t-2} + \delta_0 INCGWTH_t \\ + \delta_1 INCGWTH_{t-1} + v_t \end{aligned} \tag{9.94}$$

Does this model have lower AIC and SC values than that in (e)? Is there any evidence of serially correlated errors?

9.23 If you have not already done so, use the data in *consumptn.dat* and the sample period 1960Q4 to 2009Q4 to estimate (9.94). Given that $INCGWTH_{2010Q1} = 0.6$, $INCGWTH_{2010Q2} = 0.8$, and $INCGWTH_{2010Q3} = 0.7$, find 90% forecast intervals for consumption growth in 2010Q1, 2010Q2, and 2010Q3. Comment on these intervals.

9.24 Consider the infinite lag representation of (9.94) that we write as

$$CONGWTH_t = \alpha + \sum_{s=0}^{\infty} \beta_s INCGWTH_{t-s} + e_t$$

(a) Derive expressions that can be used to calculate the β_s from $\theta_2, \delta_0,$ and δ_1.

(b) Find estimates for the one-, two-, and three-quarter delay and interim multipliers, and the total multiplier. Interpret these estimates.

9.25 In this question we investigate the effect of wage changes on the inflation rate. Such effects can be from the demand side or the supply side. On the supply side, we expect wage increases to increase costs of production and to drive up prices. On the demand side, wage increases mean greater disposable income, and a greater demand for goods and services that also pushes up prices. Irrespective of the line of reasoning, the relationship between wage changes and inflation is likely to be a dynamic one; it takes time for wage changes to impact on inflation. To investigate this dynamic relationship, we use quarterly data on U.S. inflation (*INF*) and wage growth (*WGWTH*) from 1970Q2 to 2010Q1. These data can be found in the file *infln_wage.dat*.

(a) Graph the time series for *INF* and *WGWTH*. Include a horizontal line at the mean of each series. Do the series appear to fluctuate around a constant mean?

(b) Estimate the model $INF_t = \delta + \delta_0 WGWTH_t + v_t$. Interpret the estimate for δ_0. Check for serially correlated errors using the residual correlogram, and an *LM* test with two lagged errors. What do you conclude?

(c) Estimate the model $INF_t = \delta + \theta_1 INF_{t-1} + \delta_0 WGWTH_t + v_t$. Find estimates for the impact multiplier and the total multiplier for the effect of a change in wage growth on inflation. How do these values compare with the estimate for δ_0 from part (b)? (*Hint*: Use (9.90) with $\delta_1 = 0$ and sum the geometric progression to get the total multiplier.)

(d) Did inclusion of INF_{t-1} in the model eliminate serial correlation in the errors? Report any significant residual autocorrelations from the equation in part (c) and the results from LM tests with two and three lagged residuals.

(e) Add first INF_{t-2}, and then INF_{t-3}, to the model in part (c). In each case report the results of correlogram and LM checks for serially correlated errors.

(f) Omit INF_{t-2} from the second model estimated in part (e), and estimate the resulting model

$$INF_t = \delta + \theta_1 INF_{t-1} + \theta_3 INF_{t-3} + \delta_0 WGWTH_t + v_t \tag{9.95}$$

Why might you consider dropping INF_{t-2}? Did its omission lead to a fall in the AIC and SC? Try adding $WGWTH_{t-1}$. Does its inclusion improve the equation?

9.26 If you have not already done so, use the data in *infln_wage.dat* to estimate (9.95). Given that $WGWTH_{2010Q2} = 0.6$, $WGWTH_{2010Q3} = 0.5$, $WGWTH_{2010Q4} = 0.7$, and $WGWTH_{2011Q1} = 0.4$, find 95% forecast intervals for inflation in 2010Q2, 2010Q3, 2010Q4, and 2011Q1. Does knowing wage growth tell you much about future inflation?

9.27 Consider the infinite lag representation of (9.95) that we write as

$$INF_t = \alpha + \sum_{s=0}^{\infty} \beta_s WGWTH_{t-s} + e_t$$

(a) Derive expressions that can be used to calculate α and the β_s from $\theta_1,\ \theta_3,\ \delta,$ and δ_0.

(b) Estimate the rate of inflation when $WGWTH$ remains at $WGWTH = 0$. Use the estimates from (9.95) to test the hypothesis that the rate of inflation is zero when wage growth is zero.

(c) Estimate the rate of inflation when wage growth is constant at 0.25% per quarter.

(d) Graph the delay multipliers for lags up to 12 quarters. Comment on what this graph shows.

(e) Graph the interim multipliers for lags up to 12 quarters. Comment on what this graph shows.

(f) Suppose $WGWTH$ has been constant for a long period into the past. Then, in quarter $T + 1$ it increases by 0.2%, in quarter $T + 2$ it increases by another 0.1%, and in quarter $T + 3$ it returns to its original level. Estimate the amount by which inflation will change in quarters $T + 1,\ T + 2,\ T + 3,\ T + 4$, and $T + 5$.

Appendix 9A The Durbin-Watson Test

In Sections 9.3 and 9.4 two testing procedures for testing for autocorrelated errors, the sample correlogram and a Lagrange multiplier test, were considered. These are two large sample tests; their test statistics have their specified distributions in large samples. An alternative test, one that is exact in the sense that its distribution does not rely on a large sample approximation, is the Durbin-Watson test. It was developed in 1950 and for a long time was the standard test for $H_0 : \rho = 0$ in the AR(1) error model $e_t = \rho e_{t-1} + v_t$. It is used less frequently today because of the need to examine upper and lower bounds, as we describe below, and because its distribution no longer holds when the equation contains a lagged dependent variable.

It is assumed that the v_t are independent random errors with distribution $N(0, \sigma_v^2)$, and that the alternative hypothesis is one of positive autocorrelation. That is,

$$H_0 : \rho = 0 \qquad H_1 : \rho > 0$$

The statistic used to test H_0 against H_1 is

$$d = \frac{\sum_{t=2}^{T} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T} \hat{e}_t^2} \tag{9A.1}$$

where the $\hat{e}_t$ are the least squares residuals $\hat{e}_t = y_t - b_1 - b_2 x_t$. To see why d is a reasonable statistic for testing for autocorrelation, we expand (9A.1) as

$$\begin{aligned} d &= \frac{\sum_{t=2}^{T} \hat{e}_t^2 + \sum_{t=2}^{T} \hat{e}_{t-1}^2 - 2\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} \\ &= \frac{\sum_{t=2}^{T} \hat{e}_t^2}{\sum_{t=1}^{T} \hat{e}_t^2} + \frac{\sum_{t=2}^{T} \hat{e}_{t-1}^2}{\sum_{t=1}^{T} \hat{e}_t^2} - 2\frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2} \\ &\approx 1 + 1 - 2r_1 \end{aligned} \tag{9A.2}$$

The last line in (9A.2) holds only approximately. The first two terms differ from 1 through the exclusion of $\hat{e}_1^2$ and $\hat{e}_T^2$ from the first and second numerator summations, respectively. Thus, we have

$$d \approx 2(1 - r_1) \tag{9A.3}$$

If the estimated value of ρ is $r_1 = 0$, then the Durbin-Watson statistic $d \approx 2$, which is taken as an indication that the model errors are not autocorrelated. If the estimate of ρ happened to be $r_1 = 1$ then $d \approx 0$, and thus a low value for the Durbin-Watson statistic implies that the model errors are correlated, and $\rho > 0$.

The question we need to answer is: How close to zero does the value of the test statistic have to be before we conclude that the errors are correlated? In other words, what is a critical value d_c such that we reject H_0 when $d \leq d_c$? Determination of a critical value and a rejection region for the test requires knowledge of the probability distribution of the test statistic under the assumption that the null hypothesis, $H_0 : \rho = 0$, is true. For a 5% significance level, knowledge of the probability distribution $f(d)$ under H_0 allows us to find d_c such that $P(d \leq d_c) = 0.05$. Then, as illustrated in Figure 9A.1, we reject H_0 if $d \leq d_c$ and fail to reject H_0 if $d > d_c$. Alternatively, we can state the test procedure in terms of the p-value of the test. For this one-tail test, the p-value is given by the area under $f(d)$ to the left of

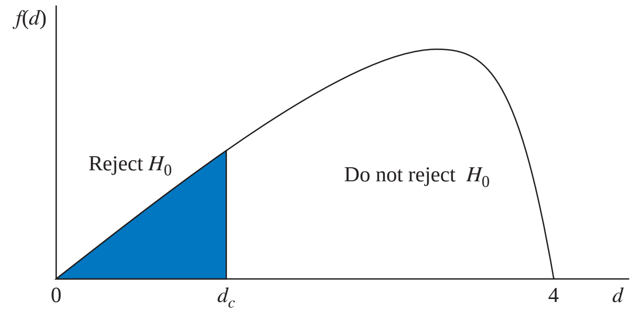

FIGURE 9A.1 Testing for positive autocorrelation.

the calculated value of d. Thus, if the p-value is less than or equal to 0.05, it follows that $d \leq d_c$, and H_0 is rejected. If the p-value is greater than 0.05, then $d > d_c$, and H_0 is accepted.

In any event, whether the test result is found by comparing d with d_c or by computing the p-value, the probability distribution $f(d)$ is required. A difficulty associated with $f(d)$, and one that we have not previously encountered when using other test statistics, is that this probability distribution depends on the values of the explanatory variables. Different sets of explanatory variables lead to different distributions for d. Because $f(d)$ depends on the values of the explanatory variables, the critical value d_c for any given problem will also depend on the values of the explanatory variables. This property means that it is impossible to tabulate critical values that can be used for every possible problem. With other test statistics, such as t, F, and χ^2, the tabulated critical values are relevant for all models.

There are two ways to overcome this problem. The first way is to use software that computes the p-value for the explanatory variables in the model under consideration. Instead of comparing the calculated d value with some tabulated values of d_c, we get our computer to calculate the p-value of the test. If this p-value is less than the specified significance level, $H_0 : \rho = 0$ is rejected, and we conclude that the errors are correlated.

In the Phillips curve example the calculated value for the Durbin-Watson statistic from the estimated equation in (9.22) is $d = 0.8873$. Is this value sufficiently close to zero (or sufficiently less than 2), to reject H_0 and conclude that autocorrelation exists? Using suitable software,[16] we find that

$$p\text{-value} = \Pr(d \leq 0.8873) = 0.0000$$

The p-value turns out to less than 10^{-6}, a value much less than a conventional 0.05 significance level; we conclude, therefore, that the equation's error is positively autocorrelated.

9A.1 The Durbin-Watson Bounds Test

In the absence of software that computes a p-value, a test known as the bounds test can be used to partially overcome the problem of not having general critical values. Durbin and Watson considered two other statistics d_L and d_U whose probability distributions do not depend on the explanatory variables and which have the property that

$$d_L < d < d_U$$

[16] The software packages SHAZAM and SAS, for example, will compute the exact Durbin-Watson p-value.

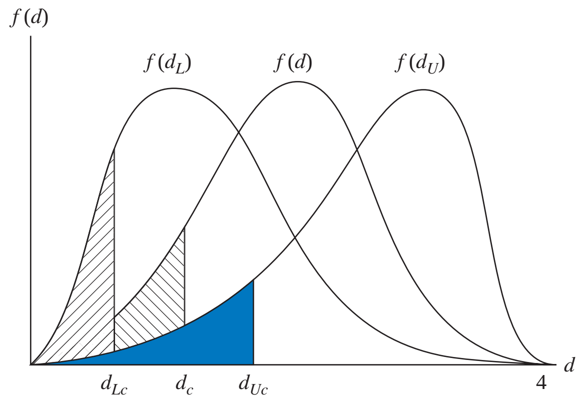

FIGURE 9A.2 Upper and lower critical value bounds for the Durbin-Watson test.

That is, irrespective of the explanatory variables in the model under consideration, d will be bounded by an upper bound d_U and a lower bound d_L. The relationship between the probability distributions $f(d_L)$, $f(d)$, and $f(d_U)$ is depicted in Figure 9A.2. Let d_{Lc} be the 5% critical value from the probability distribution for d_L. That is, d_{Lc} is such that $P(d_L < d_{Lc}) = 0.05$. Similarly, let d_{Uc} be such that $P(d_U < d_{Uc}) = 0.05$. Since the probability distributions $f(d_L)$ and $f(d_U)$ do not depend on the explanatory variables, it is possible to tabulate the critical values d_{Lc} and d_{Uc}. These values do depend on T and K, but it is possible to tabulate the alternative values for different T and K.

Thus, in Figure 9A.2 we have three critical values. The values d_{Lc} and d_{Uc} can be readily tabulated. The value d_c, the one in which we are really interested for testing purposes, cannot be found without a specialized computer program. However, it is clear from the figure that if the calculated value d is such that $d < d_{Lc}$, then it must follow that $d < d_c$, and H_0 is rejected. Also, if $d > d_{Uc}$, then it follows that $d > d_c$, and H_0 is not rejected. If it turns out that $d_{Lc} < d < d_{Uc}$, then, because we do not know the location of d_c, we cannot be sure whether to accept or reject. These considerations led Durbin and Watson to suggest the following decision rules, known collectively as the Durbin-Watson *bounds test*:

If $d < d_{Lc}$, reject $H_0 : \rho = 0$ and accept $H_1 : \rho > 0$;
if $d > d_{Uc}$, do not reject $H_0 : \rho = 0$;
if $d_{Lc} < d < d_{Uc}$, the test is inconclusive.

The presence of a range of values where no conclusion can be reached is an obvious disadvantage of the test. For this reason it is preferable to have software which can calculate the required p-value if such software is available.

The critical bounds for the Phillips curve example for $T = 90$ are[17]

$$d_{Lc} = 1.635 \quad d_{Uc} = 1.679$$

Since $d = 0.8873 < d_{Lc}$, we conclude that $d < d_c$, and hence we reject H_0; there is evidence to suggest that the errors are serially correlated.

[17] These bounds can be found from the Dubin-Watson tables on the website www.principlesofeconometrics.com.

Appendix 9B Properties of an AR(1) Error

We are interested in the mean, variance, and autocorrelations for e_t where $e_t = \rho e_{t-1} + v_t$ and the v_t are uncorrelated random errors with mean zero and variance σ_v^2. To derive the desired properties, we begin by lagging the equation $e_t = \rho e_{t-1} + v_t$ by one period, to obtain $e_{t-1} = \rho e_{t-2} + v_{t-1}$. Then, substituting e_{t-1} into the first equation yields

$$\begin{aligned} e_t &= \rho e_{t-1} + v_t \\ &= \rho(\rho e_{t-2} + v_{t-1}) + v_t \\ &= \rho^2 e_{t-2} + \rho v_{t-1} + v_t \end{aligned} \tag{9B.1}$$

Lagging $e_t = \rho e_{t-1} + v_t$ by two periods gives $e_{t-2} = \rho e_{t-3} + v_{t-2}$. Substituting this expression for e_{t-2} into (9B.1) yields

$$\begin{aligned} e_t &= \rho^2(\rho e_{t-3} + v_{t-2}) + \rho v_{t-1} + v_t \\ &= \rho^3 e_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t \end{aligned} \tag{9B.2}$$

Repeating this process k times and rearranging the order of the lagged v's yields

$$e_t = \rho^k e_{t-k} + v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \cdots + \rho^{k-1} v_{t-k+1} \tag{9B.3}$$

If we view the process as operating for a long time into the past, then we can let $k \to \infty$. This makes the first and last terms, $\rho^k e_{t-k}$ and $\rho^{k-1} v_{t-k+1}$, go to zero, because $-1 < \rho < 1$. The result is

$$e_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots \tag{9B.4}$$

The regression error e_t can be written as a weighted sum of the current and past values of the uncorrelated error v_t. This is an important result. It means that all past values of the v's have an impact on the current error e_t and that this impact feeds through into y_t through the regression equation. Notice, however, that the impact of the past v's declines the further we go into the past. The weights that are attached to the lagged v's are $\rho,\ \rho^2,\ \rho^3, \ldots$. Because $-1 < \rho < 1$, these weights decline geometrically as we consider past v's that are more distant from the current period. Eventually, they become negligible.

Equation (9B.4) can be used to find the properties of the e_t. Its mean is zero, because

$$\begin{aligned} E(e_t) &= E(v_t) + \rho E(v_{t-1}) + \rho^2 E(v_{t-2}) + \rho^3 E(v_{t-3}) + \cdots \\ &= 0 + \rho \times 0 + \rho^2 \times 0 + \rho^3 \times 0 + \cdots \\ &= 0 \end{aligned}$$

To find the variance, we write

$$\begin{aligned} \mathrm{var}(e_t) &= \mathrm{var}(v_t) + \rho^2 \mathrm{var}(v_{t-1}) + \rho^4 \mathrm{var}(v_{t-2}) + \rho^6 \mathrm{var}(v_{t-3}) + \cdots \\ &= \sigma_v^2 + \rho^2 \sigma_v^2 + \rho^4 \sigma_v^2 + \rho^6 \sigma_v^2 + \cdots \\ &= \sigma_v^2 (1 + \rho^2 + \rho^4 + \rho^6 + \cdots) \\ &= \frac{\sigma_v^2}{1 - \rho^2} \end{aligned}$$

In the above derivation zero covariance terms are ignored because the v's are uncorrelated. The result in the last line follows from rules for the sum of a geometric progression. Using shorthand notation, we have $\sigma_e^2 = \sigma_v^2/(1-\rho^2)$; the variance of e depends on that for v and the value for ρ.

To find the covariance between two e's that are one period apart, we use (9B.4) and its lag to write

$$\begin{aligned}
\operatorname{cov}(e_t, e_{t-1}) &= E(e_t e_{t-1}) \\
&= E\big[\big(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \cdots\big) \\
&\qquad \big(v_{t-1} + \rho v_{t-2} + \rho^2 v_{t-3} + \rho^3 v_{t-4} \cdots\big)\big] \\
&= \rho E\big(v_{t-1}^2\big) + \rho^3 E\big(v_{t-2}^2\big) + \rho^5 E\big(v_{t-3}^2\big) + \cdots \\
&= \rho\sigma_v^2\big(1 + \rho^2 + \rho^4 + \cdots\big) \\
&= \frac{\rho\sigma_v^2}{1-\rho^2}
\end{aligned}$$

When the second line in the above derivation is expanded, only squared terms with the same subscript are retained. Because the v's are uncorrelated, the cross-product terms with different time subscripts will have zero expectation, and are dropped from the third line. To obtain the fourth line from the third line, we have used $E(v_{t-k}^2) = \operatorname{var}(v_{t-k}) = \sigma_v^2$ for all lags k.

In a similar way, we can show that the covariance between errors that are k periods apart is

$$\operatorname{cov}(e_t, e_{t-k}) = \frac{\rho^k \sigma_v^2}{1-\rho^2} \qquad k > 0$$

Appendix 9C Generalized Least Squares Estimation

We are considering the simple regression model with AR(1) errors

$$y_t = \beta_1 + \beta_2 x_t + e_t \qquad e_t = \rho e_{t-1} + v_t$$

Our objective is to obtain the generalized least squares estimator for β_1 and β_2 by transforming the model so that it has a new uncorrelated homoskedastic error term, enabling us to apply least squares to the transformed model. To specify the transformed model we begin with (9.44), which is

$$y_t = \beta_1 + \beta_2 x_t + \rho y_{t-1} - \rho\beta_1 - \rho\beta_2 x_{t-1} + v_t \tag{9C.1}$$

and then rearrange it to give

$$y_t - \rho y_{t-1} = \beta_1(1-\rho) + \beta_2(x_t - \rho x_{t-1}) + v_t \tag{9C.2}$$

After defining the following transformed variables

$$y_t^* = y_t - \rho y_{t-1} \quad x_{t2}^* = x_t - \rho x_{t-1} \quad x_{t1}^* = 1 - \rho$$

we can rewrite (9C.2) as

$$y_t^* = x_{t1}^*\beta_1 + x_{t2}^*\beta_2 + v_t \tag{9C.3}$$

We have formed a new model with transformed variables $y_t^*, x_{t1}^*,$ and x_{t2}^* and, *importantly*, with an error term that is *not* the correlated e_t, but the uncorrelated v_t that we assumed to be distributed $(0, \sigma_v^2)$. We would expect application of least squares to (9C.3) to yield the best linear unbiased estimator for β_1 and β_2.

There are two additional problems that we need to solve, however:

1. Because lagged values of y_t and x_t had to be formed, only $(T-1)$ new observations were created by the transformation. We have values $(y_t^*, x_{t1}^*, x_{t2}^*)$ for $t = 2,3,\ldots, T$, but we have no $(y_1^*, x_{11}^*, x_{12}^*)$.
2. The value of the autoregressive parameter ρ is not known. Since y_t^*, x_{t1}^* and x_{t2}^* depend on ρ, we cannot compute these transformed observations without estimating ρ.

Considering the second problem first, we can use the sample correlation r_1 defined in (9.21) as an estimator for ρ. Alternatively, (9C.1) can be rewritten as

$$y_t - \beta_1 - \beta_2 x_t = \rho(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + v_t \tag{9C.4}$$

which is the same as $e_t = \rho e_{t-1} + v_t$. After replacing β_1 and β_2 with the least squares estimates b_1 and b_1, least squares can be applied to (9C.4) to estimate ρ.

Equations (9C.3) and (9C.4) can be estimated iteratively. That is, we use $\hat{\rho}$ from (9C.4) to estimate β_1 and β_2 from (9C.3). We then use these new estimates for β_1 and β_2 in (9C.4) to re-estimate ρ, which we then use again in (9C.3) to re-estimate β_1 and β_2, and so on. This iterative procedure is known as the Cochrane-Orcutt estimator. On convergence it is identical to the nonlinear least squares estimator described in Section 9.3.2.

What about the problem of having $(T-1)$ instead of T transformed observations? One way to solve this problem is to ignore it and to proceed with estimation on the basis of the $(T-1)$ observations. That is the strategy adopted by the estimators we have considered so far. If T is large, it is a reasonable strategy. However, if we wish to improve efficiency by including a transformation of the first observation, we need to create a transformed error that has the same variance as the errors $(v_2, v_3, \ldots, v_T)$.

The first observation in the regression model is

$$y_1 = \beta_1 + x_1\beta_2 + e_1$$

with error variance $\text{var}(e_1) = \sigma_e^2 = \sigma_v^2/(1-\rho^2)$. The transformation that yields an error variance of σ_v^2 is multiplication by $\sqrt{1-\rho^2}$. The result is

$$\sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}\beta_1 + \sqrt{1-\rho^2}x_1\beta_2 + \sqrt{1-\rho^2}e_1$$

or

$$y_1^* = x_{11}^*\beta_1 + x_{12}^*\beta_2 + e_1^* \tag{9C.5}$$

where

$$y_1^* = \sqrt{1-\rho^2}y_1 \qquad x_{11}^* = \sqrt{1-\rho^2}$$
$$x_{12}^* = \sqrt{1-\rho^2}x_1 \qquad e_1^* = \sqrt{1-\rho^2}e_1 \tag{9C.6}$$

To confirm that the variance of e_1^* is the same as that of the errors $(v_2, v_3, \ldots, v_T)$, note that

$$\text{var}(e_1^*) = (1-\rho^2)\text{var}(e_1) = (1-\rho^2)\frac{\sigma_v^2}{1-\rho^2} = \sigma_v^2$$

We also require that e_1^* be uncorrelated with $(v_2, v_3, \ldots, v_T)$. This result will hold because each of the v_t does not depend on any past values for e_t. The transformed first observation in (9C.5) can be used with the remaining transformed observations in (9C.3) to obtained generalized least squares estimates that utilize all T observations. This procedure is sometimes known as the Prais-Winsten estimator.