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. Compute the autocorrelations for a time series, graph the corresponding correlogram, and use it to test for serial correlation.
4. Explain the nature of regressions that involve lagged variables and the number of observations that are available.
5. Use autoregressive (AR) and autoregressive distributed lag (ARDL) models to compute forecasts, standard errors of forecasts, and forecast intervals.
6. Explain the assumptions required for AR and ARDL forecasting.
7. Specify and estimate ARDL models. Use serial correlation checks, significance of coefficients, and model selection criteria to choose lag lengths.
8. Test for Granger causality.
9. Use a correlogram of residuals to test for serially correlated errors.
10. Use a Lagrange multiplier test for serially correlated errors.
11. Explain the differences between time-series models for forecasting and time-series models for policy analysis.
12. Estimate and interpret the estimates from finite and infinite distributed lag models.
13. Compute HAC standard errors for least squares estimates. Explain why they are used.
14. Compute nonlinear least squares and generalized least squares estimates for a model with an AR(1) error.
15. Contrast the exogeneity assumption required for HAC standard errors with that required for estimating an AR(1) error model.
16. Compute delay, interim, and total multipliers for finite and infinite distributed lag models.

17. Test for consistency of least squares in the ARDL representation of an infinite distributed lag model.

18. Contrast the assumptions for a finite distributed lag model with those for an infinite distributed lag model.

KEYWORDS

AR(1) error
ARDL(p, q) model
autocorrelation
autoregressive distributed lags
autoregressive error
autoregressive model
correlogram
delay multiplier
distributed lag weight
dynamic models
exogeneity
finite distributed lag
forecast error
forecast intervals
forecasting
generalized least squares
geometrically declining lag
Granger causality
HAC standard errors
impact multiplier
infinite distributed lag
interim multiplier
lag length
lag operator
lagged dependent variable
LM test
moving average
multiplier analysis
nonlinear least squares
sample autocorrelations
serial correlation
standard error of forecast error
stationarity
total multiplier
$T \times R^2$ form of LM test
weak dependence

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-sectional 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 Section 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-sectional 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-sectional 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-sectional 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–evolving 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

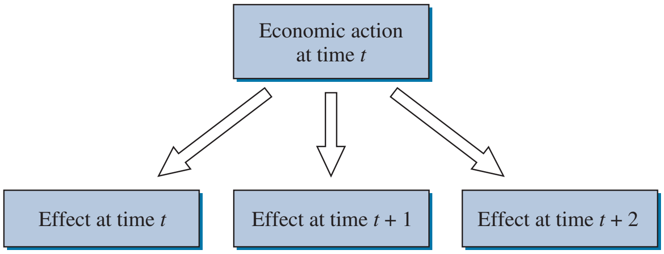

FIGURE 9.1 **The distributed lag effect.**

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 and also at times $t + 1$, $t + 2$, and so on.

EXAMPLE 9.1 | Plotting the Unemployment Rate and the GDP Growth Rate for the United States

In Figure 9.2(a) and (b), the U.S. quarterly unemployment rate, and the U.S. quarterly growth rate for gross domestic product, from 1948 quarter 1 (1948Q1) to 2016 quarter 1 (2016Q1) are graphed against time. These data can be found in the data file *usmacro*. We wish to understand how series such as these evolve over time, how current values of each data series are correlated with their past values, and how one series might be related to current and past values of another.

There are several types of models that can be used to capture the time paths of variables, their correlation structures, and their relationships with the time paths of other variables. Once a model has been selected and estimated, it may be used for **forecasting** future values or for policy analysis. We begin this chapter by describing some of the many possible time-series models and the nature of correlations between current and past values of a data series.

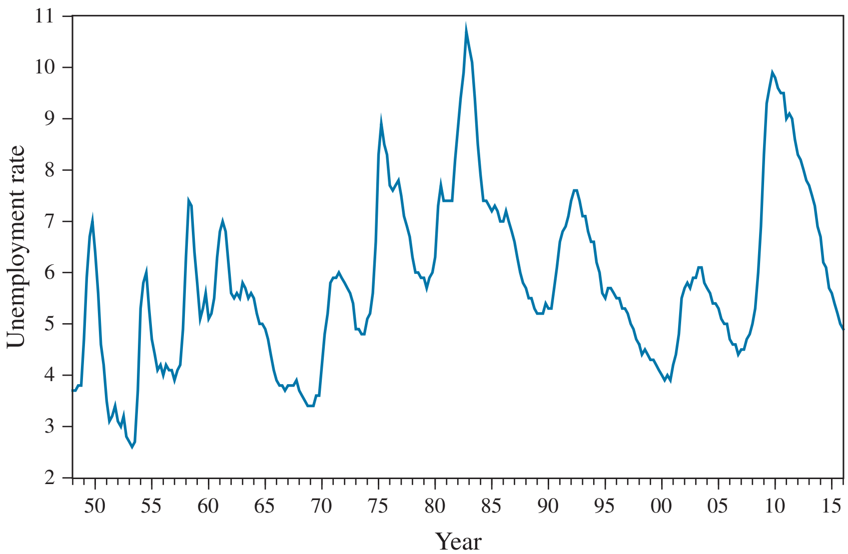

FIGURE 9.2a **U.S. Quarterly unemployment rate 1948Q1 to 2016Q1.**

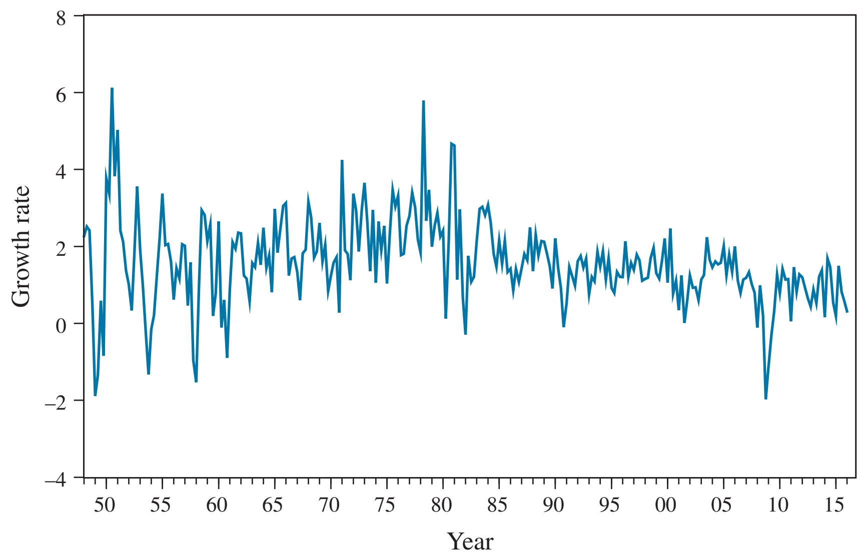

FIGURE 9.2b **U.S. GDP growth rate, 1948Q1 to 2016Q1.**

9.1.1 Modeling Dynamic Relationships

Given that time-series variables are dynamic, in the sense that their current values will be correlated with their past values, and they are related to current and past values of other variables, we need to ask how to model the dynamic nature of relationships. We can do so by introducing lagged variables into the model. These lags can take the form of lagged values of an explanatory variable $(x_{t-1}, x_{t-2}, \ldots, x_{t-q})$, lagged values of a dependent variable $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$, or lagged values of an error term $(e_{t-1}, e_{t-2}, \ldots, e_{t-s})$. In this section, we describe a number of the time-series models that arise from introducing lags of these kinds and explore the relationships between them.

Finite Distributed Lags

Suppose that the value of a variable y depends on current and past values of another variable x, up to q periods into the past. We can write this model as

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \tag{9.1}$$

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. Equations like (9.1) might say, for example, that inflation y_t depends not just on the current interest rate x_t but also on the rates in the previous q time periods $x_{t-1}, x_{t-2}, \ldots, x_{t-q}$. 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 the next q 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, equation (9.1) is called a **distributed lag model**. The coefficients β_k are sometimes known as the **lag weights**, and their sequence $\beta_0, \beta_1, \beta_2, \cdots$ is called a **lag pattern**. The model is called a **finite distributed lag** model because the effect of x on y cuts off after a finite number of periods q. Models of this kind can be used for forecasting or policy analysis. In terms of forecasting, we might be interested in using information on past interest rates to forecast future inflation. For policy analysis, a central bank might be interested in how inflation will react now and in the future to a change in the current interest rate.

The notation in (9.1) differs from what we have typically used so far. It is convenient to change the 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.

> **Remark**
>
> We use many different Greek symbols for regression parameters in this Chapter. Sometimes, it may not seem so, but our goal is clarity.

An Autoregressive Model An **autoregressive model**, or an **autoregressive process**, is one where a variable y depends on past values of itself. The general representation with p lagged values $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ is called an autoregressive model (process) of order p, abbreviated as AR(p), and is given by

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + e_t \tag{9.2}$$

For example, an AR(2) model for the unemployment rate series U in Figure 9.2(a) would be $U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + e_t$. AR models can be used to describe the time paths of variables and capture their correlations between current and past values; they are generally used for forecasting. Past values are used to forecast future values.

Autoregressive Distributed Lag Models A more general model that includes both finite distributed lag models and autoregressive models as special cases is the **autoregressive distributed lag** model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t \tag{9.3}$$

This model, with p lags of y, the current value x, and q lags of x, is abbreviated as an **ARDL(p, q) model**. 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. For example, an ARDL(2, 1) model relating the unemployment rate U to the growth rate in the economy G would be given by $U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + \delta_0 G_t + \delta_1 G_{t-1} + e_t$. ARDL models can be used for both forecasting and policy analysis. Notice that we have used "δ" with no subscript for the intercept and "δ_s" (δ with a subscript) for the coefficient of x_{t-s}. This notation is a little strange, but it avoids introducing another Greek letter for ARDL models.

Infinite Distributed Lag Models If we take equation (9.1) and assume that the impact of past, lagged x's does not cut off after q periods but goes back into the infinite past, then we have the **infinite distributed lag** (IDL) model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots + e_t \tag{9.4}$$

You might question whether values of x from a long, long time ago would still have an effect on y. You might also wonder how to decide on the cut-off point q for a finite distributed lag. One way out of this dilemma is to assume that the coefficients β_s eventually decline in magnitude with their effect becoming negligible at long lags. There are many possible lag pattern assumptions that could be made to achieve this outcome. To illustrate, consider the **geometrically declining lag** pattern

$$\beta_s = \lambda^s \beta_0, \quad 0 < \lambda < 1, \quad s = 0, 1, 2, \ldots \tag{9.5}$$

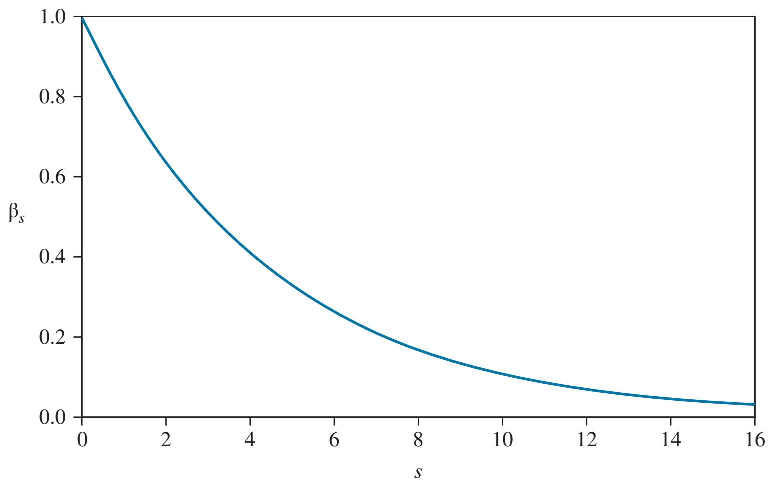

FIGURE 9.3 **Geometrically declining lag pattern.**

A graph of this lag pattern for $\beta_0 = 1$ and $\lambda = 0.8$ is displayed in Figure 9.3. Notice that, as we go back in time (s increases), β_s becomes a smaller and smaller multiple of β_0.

With the assumption in (9.5), we can write

$$y_t = \alpha + \beta_0 x_t + \lambda\beta_0 x_{t-1} + \lambda^2\beta_0 x_{t-2} + \lambda^3\beta_0 x_{t-3} + \cdots + e_t \tag{9.6}$$

Lagging this equation by one period gives the equation for y_{t-1} as

$$y_{t-1} = \alpha + \beta_0 x_{t-1} + \lambda\beta_0 x_{t-2} + \lambda^2\beta_0 x_{t-3} + \lambda^3\beta_0 x_{t-4} + \cdots + e_{t-1}$$

Multiply both sides of this equation by λ to get

$$\lambda y_{t-1} = \alpha\lambda + \lambda\beta_0 x_{t-1} + \lambda^2\beta_0 x_{t-2} + \lambda^3\beta_0 x_{t-3} + \lambda^4\beta_0 x_{t-4} + \cdots + \lambda e_{t-1} \tag{9.7}$$

Subtracting (9.7) from (9.6) gives

$$y_t - \lambda y_{t-1} = \alpha(1-\lambda) + \beta_0 x_t + e_t - \lambda e_{t-1} \tag{9.8}$$

or

$$y_t = \delta + \theta y_{t-1} + \beta_0 x_t + v_t \tag{9.9}$$

We have made the substitutions $\delta = \alpha(1-\lambda), \theta = \lambda$, and $v_t = e_t - \lambda e_{t-1}$ so that (9.9) can be recognized as an ARDL model. By making the assumption $\beta_s = \lambda^s\beta_0$, we have been able to turn the IDL model into an ARDL(1, 0) model. On the right-hand side of (9.9), there is one lag of y and the current value of x. We will see later that we can also go in the other direction. More general, ARDL(p, q) models can be turned into more flexible IDL models, providing the lagged coefficients of the IDL eventually decline and become negligible. The ARDL formulation is useful for forecasting; the IDL provides useful information for policy analysis.

An Autoregressive Error Model Another way in which lags can enter a model is through the error term. For example, if the error e_t satisfies the assumptions of an AR(1) model, it can be written as

$$e_t = \rho e_{t-1} + v_t \tag{9.10}$$

with the v_t being uncorrelated. This model means that the random error at time t is related to the random error in the previous time period plus a random component. In contrast to the AR model in (9.2), there is no intercept parameter in (9.10); it is omitted because e_t has a zero mean.

The **AR(1) error** model could be added to any of the models considered so far. To explore one of its implications, suppose that $e_t = \rho e_{t-1} + v_t$ is the error term in the model

$$y_t = \alpha + \beta_0 x_t + e_t \tag{9.11}$$

Substituting $e_t = \rho e_{t-1} + v_t$ into $y_t = \alpha + \beta_0 x_t + e_t$ yields

$$y_t = \alpha + \beta_0 x_t + \rho e_{t-1} + v_t \tag{9.12}$$

From the regression equation (9.11), the error in the previous period, time $t - 1$, can be written as

$$e_{t-1} = y_{t-1} - \alpha - \beta_0 x_{t-1} \tag{9.13}$$

Multiplying (9.13) by ρ yields

$$\rho e_{t-1} = \rho y_{t-1} - \rho\alpha - \rho\beta_0 x_{t-1} \tag{9.14}$$

Substituting (9.14) into (9.12) and rearranging yields

$$\begin{aligned} y_t &= \alpha(1-\rho) + \rho y_{t-1} + \beta_0 x_t - \rho\beta_0 x_{t-1} + v_t \\ &= \delta + \theta y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + v_t \end{aligned} \tag{9.15}$$

In the second line of (9.15), we have made the substitutions $\delta = \alpha(1-\rho)$, $\theta = \rho$ and $\beta_1 = -\rho\beta_0$ to show that it is possible to rewrite the AR(1) error model in (9.10) and (9.11) as an ARDL(1, 1) model. Equation (9.15) contains y lagged once, a current value for x, and x lagged once. However, it is a special type of ARDL model because one of its coefficients is equal to the negative product of two of the other coefficients. That is, we have the constraint, or condition, $\beta_1 = -\theta\beta_0$. **Autoregressive error** models with more lags than one can also be transformed to special cases of ARDL models.

Summary and Looking Ahead We have seen how dynamic relationships between variables can be modeled by including lags in a variety of ways. The various models are summarized in Table 9.1. There is a sense in which most of the models can be viewed as ARDL models or

TABLE 9.1 Summary of Dynamic Models for Stationary Time Series Data

Autoregressive distributed lag model, ARDL(p, q)

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t \tag{M1}$$

Finite distributed lag (FDL) model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \tag{M2}$$

Infinite distributed lag (IDL) model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots + e_t \tag{M3}$$

Autoregressive model, AR(p)

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + e_t \tag{M4}$$

Infinite distributed lag model with geometrically declining lag weights

$$\beta_s = \lambda^s \beta_0, \quad 0 < \lambda < 1, \quad y_t = \alpha(1-\lambda) + \lambda y_{t-1} + \beta_0 x_t + e_t - \lambda e_{t-1} \tag{M5}$$

Simple regression with AR(1) error

$$y_t = \alpha + \beta_0 x_t + e_t, \quad e_t = \rho e_{t-1} + v_t, \quad y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta_0 x_t - \rho\beta_0 x_{t-1} + v_t \tag{M6}$$

special cases of ARDL models. However, how we interpret and proceed with each model depends on whether the model is to be used for forecasting or policy analysis and on what assumptions are made about the error term in each model. We will examine the various scenarios as we move through the chapter. One pair of assumptions that we make throughout the chapter for all models is that the variables in the models are stationary and weakly dependent. Prior to discussing these two requirements, it is useful to introduce the concept of **autocorrelation** – also known as **serial correlation**.

9.1.2 Autocorrelations

Recall that the concepts of covariance and correlation refer to the degree of linear association between two random variables. If there is no linear association between the variables, then both the covariance and the correlation are zero. When there is some degree of linear association, correlation is the preferred measure because it is unit free and lies within the interval [−1, 1], whereas the magnitude of a covariance will depend on the units of measurement of the two variables. For two random variables, say u and v, their correlation is defined as

$$\rho_{uv} = \frac{\text{cov}(u, v)}{\sqrt{\text{var}(u)\,\text{var}(v)}} \tag{9.16}$$

If u and v are perfectly correlated, then there exist constants c and $d \neq 0$ such that $u = c + dv$, with $\rho_{uv} = 1$ when $d > 0$ and $\rho_{uv} = -1$ when $d < 0$. There is an exact linear relationship. When u and v are uncorrelated, $\rho_{uv} = \text{cov}(u, v) = 0$. Intermediate values of ρ_{uv} measure the degree of linear association.

When dealing with cross-sectional data, it is frequently reasonable to assume that each pair of observations (y_i, x_i) will be uncorrelated with other observations, a characteristic guaranteed by random sampling. In other words, $\text{cov}(y_i, y_j) = 0$ and $\text{cov}(x_i, x_j) = 0$ for $i \neq j$. With time-series data, it is unlikely that these covariances will be zero. If s is close to t, it will almost certainly be the case that $\text{cov}(y_t, y_s) \neq 0$ and $\text{cov}(x_t, x_s) \neq 0$ for $t \neq s$. Glance back at Figure 9.2(a) and (b). If unemployment is higher than average in one quarter, then, in the next quarter, it is more likely to be higher than average again, rather than lower than average. A similar statement can be made for the GDP growth rate. Changes in variables such as unemployment, output growth, inflation, and interest rates are more gradual than abrupt; their values in one period will depend on what happened in the previous period.[1] This dependence means that GDP growth now, for example, will be correlated with GDP growth in the previous period. Successive observations are likely to be correlated. Indeed, in any ARDL model where there is a linear relationship between y_t and its lags, y_t must be correlated with lagged values of itself. Correlations of this kind are called **autocorrelations**. When a variable exhibits correlation over time, we say it is **autocorrelated** or **serially correlated**. We will use these two terms interchangeably.

Let's be more precise about the definition of an autocorrelation. Consider a time series of observations on any variable, $x_1, x_2, \ldots, x_T$, with mean $E(x_t) = \mu_X$ and variance $\text{var}(x_t) = \sigma_X^2$. We assume that μ_X and σ_X^2 do not change over time. The correlation structure between x's that are observed in different time periods is described by the correlation between observations that are one period apart, the correlation between observations that are two periods apart, and so on. If we turn the formula in (9.16) into one that measures the correlation between x_t and x_{t-1}, we have

$$\rho_1 = \frac{\text{cov}(x_t, x_{t-1})}{\sqrt{\text{var}(x_t)\,\text{var}(x_{t-1})}} = \frac{\text{cov}(x_t, x_{t-1})}{\text{var}(x_t)} \tag{9.17}$$

[1] Abrupt changes can occur, particularly with financial data. Models considered in Chapter 14 can accommodate abrupt changes.

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 1**. The second equality in (9.17) holds because $\text{var}(x_t) = \text{var}(x_{t-1}) = \sigma_X^2$; we assumed that the variance does not change over time. The population autocorrelation for observations that are s periods apart is

$$\rho_s = \frac{\text{cov}(x_t, x_{t-s})}{\text{var}(x_t)} \qquad s = 1, 2, \ldots \tag{9.18}$$

Sample Autocorrelations

Population autocorrelations specified in (9.17) and (9.18) refer to a conceptual time series of observations that goes on forever, starting in the infinite past and continuing into the infinite future, $\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots$. **Sample autocorrelations** are obtained using a sample of observations for a finite time period, $x_1, x_2, \ldots, x_T$, to estimate the population autocorrelations. To estimate ρ_1 we use

$$\widehat{\text{cov}}(x_t, x_{t-1}) = \frac{1}{T-1}\sum_{t=2}^{T}(x_t - \bar{x})(x_{t-1} - \bar{x}) \quad \text{and} \quad \widehat{\text{var}}(x_t) \; \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

where $\bar{x}$ is the sample mean $\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$. The index of summation in the formula for $\widehat{\text{cov}}(x_t, x_{t-1})$ starts at $t = 2$ because we do not observe x_0. Making the substitutions, and using r_1 to denote the sample autocorrelation at lag 1, we have

$$r_1 = \frac{\sum_{t=2}^{T}(x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{9.19}$$

More generally, **the s-order sample autocorrelation** for a series x, which gives the correlation between observations that are s periods apart (the correlation between x_t and x_{t-s}), is given by

$$r_s = \frac{\sum_{t=s+1}^{T}(x_t - \bar{x})(x_{t-s} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \tag{9.20}$$

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 $(T - s)$ 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

$$r_s' = \frac{\frac{1}{T-s}\sum_{t=s+1}^{T}(x_t - \bar{x})(x_{t-s} - \bar{x})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

Another modification of (9.20) that has a similar effect is to use only $(T - s)$ observations in the denominator, so that it becomes $\sum_{t=s+1}^{T}(x_t - \bar{x})^2$. Check the computing manuals that go with this book to see which one your software uses.

Testing the Significance of an Autocorrelation

It is often useful to test whether a sample autocorrelation is significantly different from zero. That is, a test of $H_0 : \rho_s = 0$ against the alternative $H_1 : \rho_s \neq 0$. Tests of this nature are useful for constructing models and for checking whether the errors in an equation might be serially correlated. The test statistic for this test is relatively simple. When the null hypothesis $H_0: \rho_s = 0$ is true, r_s has an approximate normal distribution with mean zero and variance $1/T$. Thus, a suitable test statistic is

$$Z = \frac{r_s - 0}{\sqrt{1/T}} = \sqrt{T} r_s \overset{a}{\sim} N(0, 1) \tag{9.21}$$

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

Correlogram A useful device for assessing the significance of autocorrelations is a diagrammatic representation called 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_s will be significantly different from zero at a 5% significance level if $\sqrt{T} r_s \geq 1.96$ or if $\sqrt{T} r_s \leq -1.96$. Alternatively, we can say that r_s will be significantly different from zero if $r_s \geq 1.96/\sqrt{T}$ or $r_s \leq -1.96/\sqrt{T}$. A typical diagram for a correlogram will have bars or spikes to represent the magnitudes of the autocorrelations and approximate significant bounds drawn at $\pm 2/\sqrt{T}$, enabling the econometrician to see at a glance which correlations are significant.

EXAMPLE 9.2 | Sample Autocorrelations for Unemployment

Consider the quarterly series for the U.S. unemployment rate found in the data file *usmacro*. It runs from 1948Q1 to 2016Q1, a total of 273 observations. The first four sample autocorrelations for this series, computed from (9.20), are $r_1 = 0.967$, $r_2 = 0.898$, $r_3 = 0.811$, and $r_4 = 0.721$. The value $r_1 = 0.967$ tells us that successive values of unemployment are very highly correlated. With $r_4 = 0.721$, even observations that are four quarters apart are highly correlated. The correlogram for the unemployment rate for the first 24 lags is graphed in Figure 9.4. The heights of the bars represent the correlations. The horizontal line drawn at $2/\sqrt{173} = 0.121$ is the significance bound for positive autocorrelations. Because all the autocorrelations are positive, the negative bound of -0.121 was not included on the graph. The autocorrelations show a gradually declining pattern but remain significantly different from zero until

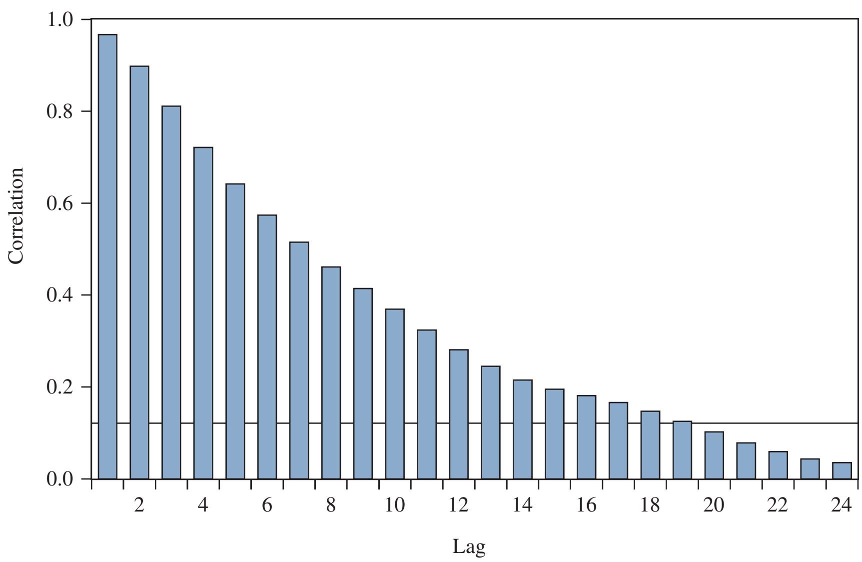

FIGURE 9.4 **Correlogram for U.S. quarterly unemployment rate.**

lag 19, beyond which they are not statistically significant. As the chapter evolves, we will discover that estimates of autocorrelations are important for model construction and checking whether one of our assumptions is violated.

Your software might not produce a correlogram that is exactly the same as Figure 9.4. 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 its significance bounds might be slightly different than ours. Be prepared! Learn to isolate and focus on the information corresponding to that in Figure 9.4 and do not be disturbed if the output is slightly but not substantially different. If the significance bounds are slightly different, it is because they use a different refinement of the large sample approximation $\sqrt{T}r_s \overset{a}{\sim} N(0, 1)$.

EXAMPLE 9.3 | Sample Autocorrelations for GDP Growth Rate

As a second example of sample autocorrelations and the associated correlogram, we consider quarterly data for the U.S. GDP growth rate that can also be found in the data file *usmacro*. In this case, the first four sample autocorrelations are $r_1 = 0.507$, $r_2 = 0.369$, $r_3 = 0.149$, and $r_4 = 0.085$; the correlogram for up to 48 lags is presented in Figure 9.5. These correlations are much smaller than those for the unemployment series, but there is a seemingly strange pattern where the correlations, although not large, oscillate between significance and insignificance at longer lags. This is a complex structure, perhaps attributable to the business cycle.

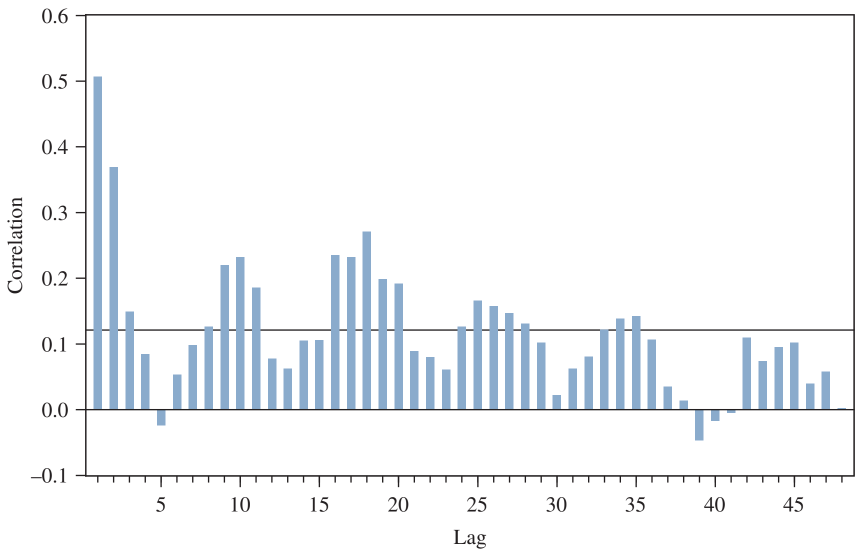

FIGURE 9.5 **Correlogram for growth rate in U.S. GDP.**

9.2 Stationarity and Weak Dependence

A critical assumption that is maintained throughout this chapter is that the variables in our equations are **stationary**. Stationary variables have means and variances that do not change over time and autocorrelations that depend only on how far apart the observations are in time, not on a particular point in time. Specifically, the autocorrelations in (9.18) depend on the time between the periods s, but not the actual point in time t. Implicit in the discussion in Section 9.1.2 was that x_t is stationary. Its mean μ_X, variance σ_X^2, and autocorrelations ρ_s were assumed not to be different for different t. In Examples 9.2 and 9.3, autocorrelations for the unemployment and growth rates were calculated under the assumption that both are stationary. Saying that a series is stationary implies that, if we took different subsets of observations corresponding to different windows of time, and used them for estimation, we would be estimating the same population quantities, the same mean μ, the same variance σ^2, and the same autocorrelations $\rho_1, \rho_2, \rho_3, \cdots$.

The first task when estimating a relationship with time-series data is to plot the observations on the variables, as we did in Figure 9.2(a) and (b), to gain an appreciation of the nature of your data and to see if there is evidence of nonstationarity. In addition, formal tests known as **unit root tests** can be used to detect nonstationarity. These tests and strategies for estimation with nonstationary variables are considered in Chapter 12. Because checking for nonstationarity is an essential first step, some readers may wish to temporarily jump forward to unit root testing in Chapter 12 before returning to our coverage of estimation and forecasting with stationary variables. For the moment, we note that a stationary variable is one that is not explosive, nor is it trending, and nor does it wander aimlessly without returning to its mean. These features can be illustrated with some graphs. Figure 9.6(a–c) contains graphs of simulated 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 9.6(a) is considered stationary because it tends to fluctuate around a constant mean without wandering or trending. On the other hand, x and z that appear in Figure 9.6(b) and (c) possess characteristics of nonstationary variables. In Figure 9.6(b), x tends to wander or is "slow turning," while z in Figure 9.4(c) is trending. These concepts will be defined more precisely in Chapter 12. At the present time, 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."

In addition to assuming that the variables are stationary, in this chapter we also assume they are **weakly dependent**. **Weak dependence** implies that, as $s \to \infty$ (observations get further and further apart in time), they become almost independent. For s large enough, the autocorrelations ρ_s become negligible. When using correlated time-series variables, weak dependence is needed for the least squares estimator to have desirable large sample properties. Typically, stationary variables have weak dependence. However, there are rare exceptions.

EXAMPLE 9.4 | Are the Unemployment and Growth Rate Series Stationary and Weakly Dependent?

A formal checking of the unemployment and growth rate series for **stationarity** is deferred until unit root tests are introduced in Chapter 12. It is useful, however, to see what tentative conclusions might be drawn from the plots and correlograms of the two series. An examination of the plot for unemployment in Figure 9.2(a) suggests that it has characteristics that make it more similar to Figure 9.6(b) than to Figure 9.6(a). Thus, on the basis of the plot alone, one might be inclined to conclude the unemployment rate is nonstationary. It turns out that a unit root test rejects a null hypothesis of nonstationarity, suggesting that the series can be treated as stationary, but its very high autocorrelations have led to the wandering characteristics exhibited in Figure 9.2(a). Do we have evidence to suggest that the series is weakly dependent? The answer is yes. The autocorrelations in the correlogram in Figure 9.4 are becoming smaller and smaller at longer lags and eventually die out to $r_{24} = 0.035$. Had we considered lags beyond 24, we would find $r_{36} = 0.008$.

Turning to the GDP growth series, we note that its plot in Figure 9.2b has characteristics similar to those of Figure 9.6(a), enabling us to tentatively conclude that it is stationary. GDP growth has ups and downs from one quarter to the next, but it does not keep going up or down for long periods; it returns to the middle, or mean, after a short time. Its correlogram in Figure 9.5 has some significant correlations at long lags, but they are not large and, when autocorrelations beyond those displayed in Figure 9.5 are examined, they die out very quickly, leading us to conclude the series is weakly dependent.

Knowing the unemployment and growth rates are stationary and weakly dependent means that we can proceed to use them for the examples in this chapter devoted to time-series regression models with stationary variables. With the exception of a special case known as cointegration—considered in Chapter 12—variables in time-series regressions must be stationary and weakly dependent for the least squares estimator to be consistent.

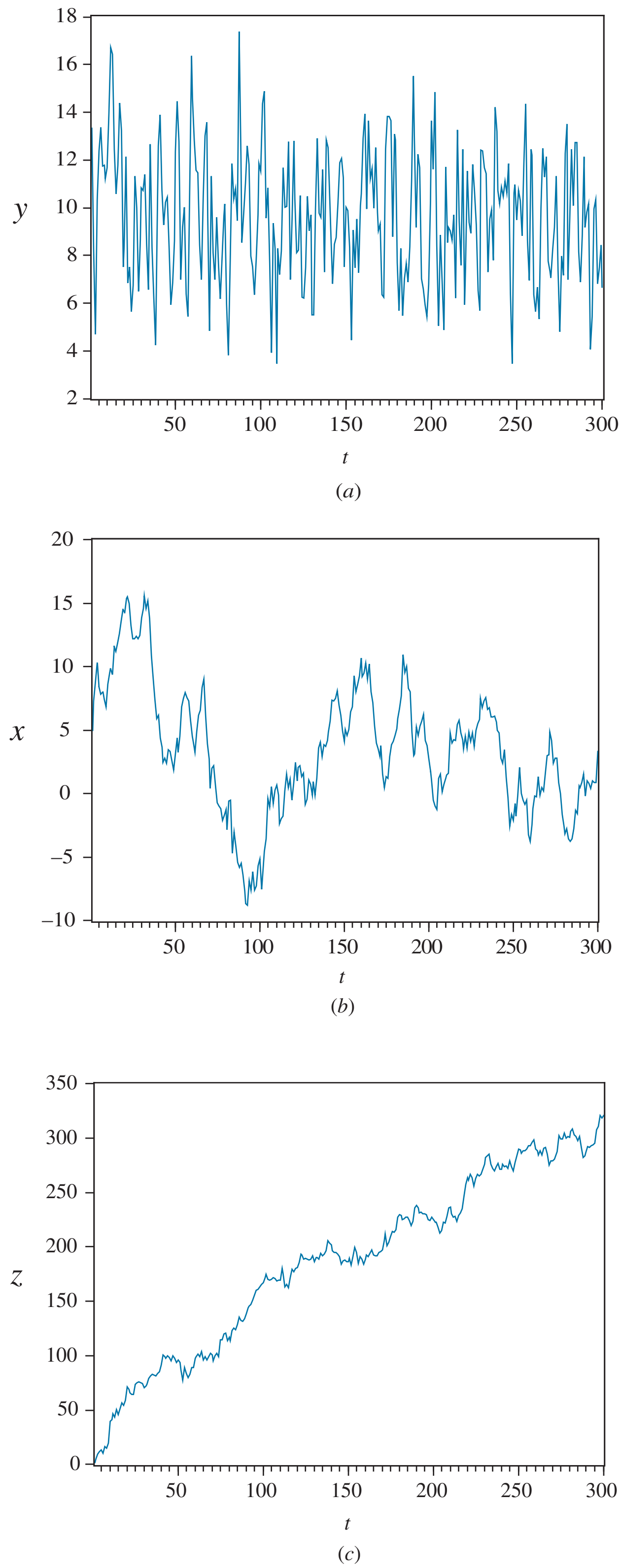

FIGURE 9.6 **(a) Time series of a stationary variable; (b) time series of a nonstationary variable that is “slow-turning” or “wandering”; (c) time series of a nonstationary variable that “trends.”**

9.3 Forecasting

The forecasting of 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.[2] In this section, we consider forecasting using two different models, an AR model, and an ARDL model. Our focus is on **short-term forecasting**, typically up to three periods into the future.

To introduce the forecasting problem within the context of an ARDL model, suppose that we are given the following ARDL(2,2) model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \delta_1 x_{t-1} + \delta_2 x_{t-2} + e_t \tag{9.22}$$

Criteria for choosing the numbers of lags for y and x will be discussed in Sections 9.3.3 and 9.4. For the moment, we use two lags of each to describe the essential features of the forecasting problem. A quick comparison of (9.22) with (9.3) reveals a slight difference: the term $\delta_0 x_t$ has been omitted from (9.22). To appreciate why, suppose that we have the sample observations $\{(y_t, x_t), t = 1, 2, \ldots, T\}$ and that we wish to forecast y_{T+1} which, from (9.22), is given by

$$y_{T+1} = \delta + \theta_1 y_T + \theta_2 y_{T-1} + \delta_1 x_T + \delta_2 x_{T-1} + e_{T+1} \tag{9.23}$$

Including $\delta_0 x_t$ in (9.22) would mean including $\delta_0 x_{T+1}$ in (9.23). If the future value x_{T+1} were known, then its inclusion is desirable, but the more likely situation is that both y_{T+1} and x_{T+1} will not be observed at time T when the forecast is made. Thus, dropping x_t in (9.22) is a more practical choice.

Define the information set of all current and past observations on y and x at time t as

$$I_t = \{y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots\} \tag{9.24}$$

Assuming that we are standing at the end of the sample period, having observed y_T and x_T, the one-period ahead forecasting problem is to find a forecast $\hat{y}_{T+1}$ conditional on, or given, the information at time T, $I_T = \{y_T, y_{T-1}, \ldots, x_T, x_{T-1}, \ldots\}$. If the parameters $(\delta, \theta_1, \theta_2, \delta_1, \delta_2)$ are known, the best forecast in the sense that it minimizes conditional mean-squared **forecast error** $E\left[\left(\hat{y}_{T+1} - y_{T+1}\right)^2 \middle| I_T\right]$ is the conditional expectation $\hat{y}_{T+1} = E(y_{T+1}|I_T)$. We investigate what this implies for the ARDL(2, 2) model in (9.23) and later discuss estimation of the parameters. If we believe that only two lags of y and two lags of x are relevant—they provide the best forecast—we are assuming that

$$\begin{aligned} E(y_{T+1}|I_T) &= E(y_{T+1}|y_T, y_{T-1}, x_T, x_{T-1}) \\ &= \delta + \theta_1 y_T + \theta_2 y_{T-1} + \delta_1 x_T + \delta_2 x_{T-1} \end{aligned} \tag{9.25}$$

Notice the difference between the two conditional expectations: $E(y_{T+1}|I_T)$ conditions on all past observations; $E(y_{T+1}|y_T, y_{T-1}, x_T, x_{T-1})$ conditions on only the two most recent observations. By employing an ARDL(2, 2) model, we are assuming that, for forecasting y_{T+1}, observations from more than two periods in the past do not convey any extra information relative to that contained in the most recent two observations. In addition, for the result in (9.25) to hold, we require

$$E(e_{T+1}|I_T) = 0 \tag{9.26}$$

[2]A comprehensive but relatively advanced treatment is Graham Elliott and Allan Timmermann, *Economic Forecasting*, 2016, Princeton University Press.

For two-period ahead and three-period ahead forecasts, the best forecasts are given respectively by

$$\hat{y}_{T+2} = E(y_{T+2}|I_T) = \delta + \theta_1 E(y_{T+1}|I_T) + \theta_2 y_T + \delta_1 E(x_{T+1}|I_T) + \delta_2 x_T$$
$$\hat{y}_{T+3} = E(y_{T+3}|I_T) = \delta + \theta_1 E(y_{T+2}|I_T) + \theta_2 E(y_{T+1}|I_T) + \delta_1 E(x_{T+2}|I_T) + \delta_2 E(x_{T+1}|I_T)$$

Notice the extra requirements for these two forecasts. We need to know $E(y_{T+2}|I_T)$, $E(y_{T+1}|I_T)$, $E(x_{T+2}|I_T)$ and $E(x_{T+1}|I_T)$. We have estimates of $E(y_{T+2}|I_T)$ and $E(y_{T+1}|I_T)$ readily available from previous periods' forecasts, but $E(x_{T+2}|I_T)$ and $E(x_{T+1}|I_T)$ require extra information. This information can come from independent forecasts or we might be interested in what-if type questions such as if the next two future values of x are $\hat{x}_{T+1}$ and $\hat{x}_{T+2}$, what will be the point and interval forecasts for y_{T+2} and y_{T+3}? If the model is a pure autoregressive one without an x-component, this issue does not arise. In what follows we first consider an example using a pure AR model, and then one with one lagged x. These are both special cases of (9.22). We defer discussion of (9.26) and other assumptions until after the examples.

EXAMPLE 9.5 | Forecasting Unemployment with an AR(2) Model

To demonstrate how to use an AR model for forecasting, we consider the following AR(2) model for forecasting the U.S. unemployment rate U

$$U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + e_t \tag{9.27}$$

The aim is to use observations up to and including 2016Q1 to forecast unemployment in the next three quarters: 2016Q2, 2016Q3, and 2016Q4. The information set at time t is $I_t = \{U_t, U_{t-1}, \ldots\}$. At the time we have observed 2016Q1, it is $I_{2016Q1} = \{U_{2016Q1}, U_{2015Q4}, \ldots\}$. We assume that (9.26) holds which, in general terms for any time period, can be written as $E(e_t|I_{t-1}) = 0$. Past values of unemployment cannot be used to forecast the error in the current period. With this set up, we can write expressions for forecasts for the remainder of 2016 as

$$\hat{U}_{2016Q2} = E(U_{2016Q2}|I_{2016Q1}) = \delta + \theta_1 U_{2016Q1} + \theta_2 U_{2015Q4} \tag{9.28}$$

$$\begin{aligned}\hat{U}_{2016Q3} &= E(U_{2016Q3}|I_{2016Q1})\\ &= \delta + \theta_1 E(U_{2016Q2}|I_{2016Q1}) + \theta_2 U_{2016Q1}\end{aligned} \tag{9.29}$$

$$\begin{aligned}\hat{U}_{2016Q4} &= E(U_{2016Q4}|I_{2016Q1})\\ &= \delta + \theta_1 E(U_{2016Q3}|I_{2016Q1}) + \theta_2 E(U_{2016Q2}|I_{2016Q1})\end{aligned} \tag{9.30}$$

Because these expressions all depend on the unknown parameters (δ, θ_1, θ_2), before we can proceed we need to estimate them. We digress for a moment to consider estimation of the AR(2) model.

OLS Estimation of the AR(2) Model for Unemployment

The assumption $E(e_t|I_{t-1}) = 0$ is sufficient for the OLS estimator for $(\delta, \theta_1, \theta_2)$ to be consistent. The OLS estimator will not be unbiased, but consistency gives it a large-sample justification. Assuming that $E(e_t|I_{t-1}) = 0$ is weaker than the strict **exogeneity** assumption. In the general ARDL model, it implies $\text{cov}(e_t, y_{t-s}) = 0$ and $\text{cov}(e_t, x_{t-s}) = 0$ for all $s > 0$ but it does not preclude future values y_{t+s} and x_{t+s}, $s > 0$, from being correlated with e_t. The model in (9.27) can be treated in the same way as the multiple regression model in Chapters 5 and 6, with $U_{t-1} = x_{t1}$ and $U_{t-2} = x_{t2}$. The two lags of the "dependent variable" can be treated as two different explanatory variables. One difference is that the two lags cause us to lose two observations. Instead of having $T = 273$ observations for estimation, only $T - 2 = 271$ are available. From a practical standpoint, this modification is not a concern; the software that you are using will make the necessary adjustments. It is nevertheless useful to fully appreciate how the lagged variables are defined and how their observations enter the estimation procedure. Table 9.2 contains the observations as separate variables in the form they would appear in a spreadsheet. Notice how the observations are lagged and how we lose one observation when U_{t-1} is formed, and two observations when U_{t-2} is formed.

TABLE 9.2 Spreadsheet of Observations for AR(2) Model

t	Quarter	U_t	U_{t-1}	U_{t-2}
1	1948Q1	3.7	•	•
2	1948Q2	3.7	3.7	•
3	1948Q3	3.8	3.7	3.7
4	1948Q4	3.8	3.8	3.7
5	1949Q1	4.7	3.8	3.8
⋮	⋮	⋮	⋮	⋮
271	2015Q3	5.2	5.4	5.6
272	2015Q4	5.0	5.2	5.4
273	2016Q1	4.9	5.0	5.2

Using the observations in Table 9.2 to find OLS estimates of the model in equation (9.27) yields

$$\begin{aligned} \hat{U}_t &= 0.2885 + 1.6128U_{t-1} - 0.6621U_{t-2} \qquad \hat{\sigma} = 0.2947 \\ \text{(se)} &\quad (0.0666) \quad (0.0457) \qquad (0.0456) \end{aligned} \tag{9.31}$$

The standard errors in this equation are the conventional least squares standard errors introduced in Chapters 2 and 5. These standard errors and the estimate $\hat{\sigma} = 0.2947$ will be valid with the conditional homoskedasticity assumption $\text{var}\left(e_t|U_{t-1}, U_{t-2}\right) = \sigma^2$. In addition, in large samples, the usual t- and F-statistics are valid for testing hypotheses or constructing interval estimates for $\left(\delta, \theta_1, \theta_2\right)$. You might wonder whether we need an assumption corresponding to the one made in Chapters 2 and 5, that the errors are serially uncorrelated. It can be shown that one of the assumptions that has already been made, $E\left(U_t|I_{t-1}\right) = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2}$, implies that the errors are uncorrelated.[3]

Unemployment Forecasts Having estimated the AR(2) model, we are now in a position to use it for forecasting. Recognizing that the unemployment rates for the two most recent quarters are $U_{2016Q1} = 4.9$ and $U_{2015Q4} = 5$, the forecast for U_{2016Q2} obtained using (9.28) and the estimates in (9.31) is[4]

$$\begin{aligned} \hat{U}_{2016Q2} &= \hat{\delta} + \hat{\theta}_1 U_{2016Q1} + \hat{\theta}_2 U_{2015Q4} \\ &= 0.28852 + 1.61282 \times 4.9 - 0.66209 \times 5 \\ &= 4.8809 \end{aligned} \tag{9.32}$$

Moving to the forecast for two quarters ahead, we have

$$\begin{aligned} \hat{U}_{2016Q3} &= \hat{\delta} + \hat{\theta}_1 \hat{U}_{2016Q2} + \hat{\theta}_2 U_{2016Q1} \\ &= 0.28852 + 1.61282 \times 4.8809 - 0.66209 \times 4.9 \\ &= 4.9163 \end{aligned} \tag{9.33}$$

There is an important difference in the way the forecasts $\hat{U}_{2016Q2}$ and $\hat{U}_{2016Q3}$ are obtained. It is possible to calculate $\hat{U}_{2016Q2}$ using only past observations on U. However, U_{2016Q3} depends on U_{2016Q2}, which is unobserved at time 2016Q1. To overcome this problem, we replace U_{2016Q2} by

[3]See Exercise 9.3 for an example where autocorrelated errors imply an extra lag of the dependent variable should be included.

[4]We carry the coefficient estimates to five decimal places to reduce rounding error.

its forecast $\hat{U}_{2016Q2}$ on the right side of equation (9.33). For forecasting U_{2016Q4}, forecasts for both U_{2016Q3} and U_{2016Q2} are needed on the right side of the equation. Specifically,

$$\begin{aligned}\hat{U}_{2016Q4} &= \hat{\delta} + \hat{\theta}_1 \hat{U}_{2016Q3} + \hat{\theta}_2 \hat{U}_{2016Q2} \\ &= 0.28852 + 1.61282 \times 4.9163 - 0.66209 \times 4.8809 \\ &= 4.986 \end{aligned} \tag{9.34}$$

The forecast unemployment rates for 2016Q2, 2016Q3, and 2016Q4 are approximately 4.88%, 4.92%, and 4.99%, respectively. By the time this book is published, we will be able to compare these forecasts with what actually happened!

9.3.1 Forecast Intervals and Standard Errors

We are typically interested in not just point forecasts but also interval forecasts that give a likely range in which a future value could fall and indicate the reliability of a point forecast. To investigate how to construct a forecast interval, we return to the more general ARDL(2, 2) model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \delta_1 x_{t-1} + \delta_2 x_{t-2} + e_t$$

and examine the forecast errors for one-period, two-period, and three-period ahead forecasts. The one-period ahead forecast error f_1 is given by

$$\begin{aligned} f_1 = y_{T+1} - \hat{y}_{T+1} = \left(\delta - \hat{\delta}\right) + \left(\theta_1 - \hat{\theta}_1\right) y_T + \left(\theta_2 - \hat{\theta}_2\right) y_{T-1} + \left(\delta_1 - \hat{\delta}_1\right) x_{T-1} \\ + \left(\delta_2 - \hat{\delta}_2\right) x_{T-2} + e_{T+1} \end{aligned} \tag{9.35}$$

where $\left(\hat{\delta}, \hat{\theta}_1, \hat{\theta}_2, \hat{\delta}_1, \hat{\delta}_2\right)$ are the least squares estimates. The difference between the forecast $\hat{y}_{T+1}$ and the corresponding realized value y_{T+1} depends on the differences between the actual coefficients and the estimated coefficients and on the value of the random error e_{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 in time-series forecasting because the variance of the random error is typically large relative to the variances of the estimated coefficients, and the resulting estimator for the forecast error variance retains the property of consistency. This means that we can write the forecast error for one quarter ahead as

$$f_1 = e_{T+1} \tag{9.36}$$

For two periods ahead, the forecast error gets more complicated. In this case, ignoring sampling error from estimating the coefficients, we will be using

$$\hat{y}_{T+2} = \delta + \theta_1 \hat{y}_{T+1} + \theta_2 y_T + \delta_1 \hat{x}_{T+1} + \delta_2 x_T \tag{9.37}$$

to forecast

$$y_{T+2} = \delta + \theta_1 y_{T+1} + \theta_2 y_T + \delta_1 x_{T+1} + \delta_2 x_T + e_{T+2} \tag{9.38}$$

In (9.37), $\hat{y}_{T+1}$ comes from the one-period ahead forecast, but a value for $\hat{x}_{T+1}$ needs to be obtained from elsewhere. To forecast two periods ahead, we will also need $\hat{x}_{T+2}$. These values may come from their own forecasting model, or they might be set by the forecaster to answer what-if type questions. We will assume that these values are given, $\hat{x}_{T+1} = x_{T+1}$ and $\hat{x}_{T+2} = x_{T+2}$, or, alternatively, that we are asking what-if type questions so that we can assume that there is no error from predicting future values of x. Given these assumptions, the two-period ahead forecast error is

$$f_2 = \theta_1 \left(y_{T+1} - \hat{y}_{T+1}\right) + e_{T+2} = \theta_1 f_1 + e_{T+2} = \theta_1 e_{T+1} + e_{T+2} \tag{9.39}$$

For three periods ahead, the error can be shown to be

$$f_3 = \theta_1 f_2 + \theta_2 f_1 + e_{T+3} = \left(\theta_1^2 + \theta_2\right) e_{T+1} + \theta_1 e_{T+2} + e_{T+3} \tag{9.40}$$

Expressing the forecast errors in terms of the e_t's is convenient for deriving expressions for the forecast error variances. With the assumptions $E(e_t|I_{t-1}) = 0$ and $\text{var}(e_t|y_{t-1}, y_{t-2}, x_{t-1}, x_{t-2}) = \sigma^2$, equations (9.36), (9.39), and (9.40) can be used to show that

$$\sigma_{f1}^2 = \text{var}(f_1|I_T) = \sigma^2$$
$$\sigma_{f2}^2 = \text{var}(f_2|I_T) = \sigma^2(1 + \theta_1^2)$$
$$\sigma_{f3}^2 = \text{var}(f_3|I_T) = \sigma^2\left[(\theta_1^2 + \theta_2)^2 + \theta_1^2 + 1\right] \quad (9.41)$$

The standard errors of the forecast errors are obtained by replacing the unknown parameters in (9.41) by their estimates and then taking the square root. Denoting these standard errors by $\hat{\sigma}_{f1}$, $\hat{\sigma}_{f2}$, and $\hat{\sigma}_{f3}$, $100(1-\alpha)\%$ **forecast intervals** are given by $\hat{y}_{T+j} \pm t_{(1-\alpha/2, T-7)}\hat{\sigma}_{fj}$, $j = 1, 2, 3$. The degrees of freedom for the t-distribution are $(T - p - q - 1) - 2 = T - 7$ because five coefficients have been estimated and the two lags have led to a loss of two observations.[5]

EXAMPLE 9.6 | Forecast Intervals for Unemployment from the AR(2) Model

Using the forecast-error variances in (9.41), the estimates in (9.31) and $t_{(0.975, 268)} = 1.9689$, we can compute the forecast standard errors and 95% forecast intervals presented in Table 9.3. Notice how the forecast standard errors and the widths of the intervals increase as we forecast further into the future, reflecting the extra uncertainty from doing so. It is much harder to be precise about forecasts further into the future. This idea was introduced in Figure 4.2.

TABLE 9.3 Forecasts and Forecast Intervals for Unemployment from AR(2) Model

Quarter	Forecast $\hat{U}_{T+j}$	Standard Error of Forecast Error ($\hat{\sigma}_{fj}$)	Forecast Interval $(\hat{U}_{T+j} \pm 1.9689 \times \hat{\sigma}_{fj})$
2016Q2 ($j = 1$)	4.881	0.2947	(4.301, 5.461)
2016Q3 ($j = 2$)	4.916	0.5593	(3.815, 6.017)
2016Q4 ($j = 3$)	4.986	0.7996	(3.412, 6.560)

EXAMPLE 9.7 | Forecasting Unemployment with an ARDL(2, 1) Model

In this example, we include a lagged value of the growth rate of GDP (G) to see if its inclusion improves the precision of our forecasts. We would expect a high growth rate to lead to less unemployment and a slowdown in the economy to create more unemployment. The least squares estimated model is

$$\hat{U}_t = \underset{(0.0723)}{0.3616} + \underset{(0.0556)}{1.5331}U_{t-1} - \underset{(0.0556)}{0.5818}U_{t-2} - \underset{(0.01949)}{0.04824}G_{t-1} \quad \text{(se)}$$
$$\hat{\sigma} = 0.2919 \quad (9.42)$$

Apart from the need to supply future values of G necessary for forecasting more than one quarter into the future, 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. In Exercise 9.16, you are invited to verify the values reported in Table 9.4. For the forecasts for

[5]The large sample distribution theory upon which this forecast interval is based uses a normal distribution rather than a t-distribution. Thus, the interval $\hat{y}_{T+j} \pm z_{1-\alpha/2}\hat{\sigma}_{fj}$ is also used. The t-distribution is frequently chosen in practice to be more conservative.

2016Q3 and 2016Q4, we assumed that $G_{2016Q2} = 0.869$ and $G_{2016Q3} = 1.069$. Comparing the forecasts in Tables 9.3 and 9.4, we find that including the lagged growth rate has increased the point forecasts for unemployment and reduced slightly the standard errors of the forecasts. The main source of the larger point forecasts appears to be the increase in the estimate of the intercept δ from 0.2885 to 0.3616. In addition, although the values $G_{2016Q2} = 0.869$ and $G_{2016Q3} = 1.069$ assume an improved growth rate relative to $G_{2016Q1} = 0.310$, they are still below the sample average growth rate of $\overline{G} = 1.575$.

TABLE 9.4 **Forecasts and Forecast Intervals for Unemployment from ARDL(2, 1) Model**

Quarter	Forecast $\hat{U}_{T+j}$	Standard Error of Forecast Error $(\hat{\sigma}_{fj})$	Forecast Interval $\left(\hat{U}_{T+j} \pm 1.9689 \times \hat{\sigma}_{fj}\right)$
2016Q2 ($j = 1$)	4.950	0.2919	(4.375, 5.525)
2016Q3 ($j = 2$)	5.058	0.5343	(4.006, 6.110)
2016Q4 ($j = 3$)	5.184	0.7430	(3.721, 6.647)

We have considered forecasting with both AR and ARDL models. It remains to point out that forecasting with a finite distributed lag model with no AR component can be carried out within the same framework as forecasting in the linear regression model that we considered in Section 6.4. Instead of the right-hand-side variables being a number of different x's, they comprise a number of lags on the same x.

9.3.2 Assumptions for Forecasting

Throughout this section, we have alluded to the various assumptions that ensure an ARDL model can be estimated consistently and used for forecasting. A summary of these assumptions and some of their implications follows.

F1: The time series y and x are stationary and weakly dependent. How to test this assumption and how to model time series that violate the assumption are considered in Chapter 12.

F2: The conditional expectation $E(y_t|I_{t-1})$ is a linear function of a finite number of lags of y and x. That is,

$$E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} \tag{9.43}$$

where $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\}$ is defined as the information set at time $t-1$ and represents all past observations at time t. There are a number of things implied by this assumption.

1. Lags of y beyond y_{t-p} and lags of x beyond x_{t-q} do not contribute to the conditional expectation; they cannot improve the forecast of y_t.
2. The error term e_t in the ARDL model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t$$

is such that $E(e_t|I_{t-1}) = 0$.
3. Let $\mathbf{z}_t = (1, y_{t-1}, \ldots, y_{t-p}, x_{t-1}, \ldots x_{t-q})$ denote all right-hand side variables in the ARDL model at time t. The e_t are not serially correlated in the sense that $E(e_t e_s|\mathbf{z}_t, \mathbf{z}_s) = 0$ for $t \neq s$. If the e_t were serially correlated, then at least one

more lag of y should appear in $E(y_t|I_{t-1})$. To gain an intuitive appreciation of why this is so, consider the AR(1) model $y_t = \delta + \theta_1 y_{t-1} + e_t$. Correlation between e_t and e_{t-1} implies that we can write $E(e_t|I_{t-1}) = \rho e_{t-1}$, from which we obtain $E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \rho e_{t-1}$. From the original model, $e_{t-1} = y_{t-1} - \delta - \theta_1 y_{t-2}$, and so

$$E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \rho(y_{t-1} - \delta - \theta_1 y_{t-2})$$
$$= \delta(1-\rho) + (\theta_1 + \rho)\, y_{t-1} - \rho\theta_1 y_{t-2}$$

4. The assumption $E(e_t|I_{t-1}) = 0$ does not preclude feedback from a past error e_{t-j} $(j > 0)$ to current and future values of x. If x is a policy variable whose setting reacts to past values of e and y, the least squares estimator is still consistent and the conditional expectation remains the best forecast. Correlation between e_t and past values of x is excluded, however. If e_t was correlated with x_{t-1} (say), then $E(e_t|I_{t-1}) \neq 0$.

F3: The errors are conditionally homoskedastic, $\text{var}(e_t|\mathbf{z}_t) = \sigma^2$. This assumption is needed for the traditional least squares standard errors to be valid and to compute the forecast standard errors.

9.3.3 Selecting Lag Lengths

So far in our description of an ARDL model and how it can be used for forecasting, we have taken the **lag lengths** p and q as given. A critical assumption to ensure that we had the best forecast in a minimum mean-squared-error sense was that no lags beyond those included in the model contained extra information that could improve the forecast. Technically, this assumption was equivalent to $E(e_t|I_{t-1}) = 0$ where e_t is the equation error term, and $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\}$ is the set of information prior to period t. A natural question that now arises is: How many lags of y and x should be included? Specifically, in terms of the ARDL(p, q) model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t$$

how do we decide on p and q? There are a number of different criteria that can be used. Because they all do not necessarily lead to the same choice, there is a degree of subjective judgment that must be exercised. It is an area in which econometrics is an art as well as a science.

We can explain three criteria relatively quickly. One is to extend the lag lengths for y and x as long as their estimated coefficients are significantly different from zero. A second is to choose p and q to minimize either the AIC or the SC variable selection criterion. And a third is to evaluate the out-of-sample forecasting performance of each (p, q) combination using a hold-out sample. Testing significance was introduced in Chapter 3 and has been used extensively since. The second and third criteria were discussed in Section 6.4.1. In the remainder of this section, we use the unemployment equation to illustrate how the SC can be used to choose lag lengths.[6] A fourth way of of deciding on p and q is to check for serial correlation in the error term. Since $E(e_t|I_{t-1}) = 0$ implies that the lag lengths p and q are sufficient and the errors are not serially correlated, the presence of serial correlation is an indication we have insufficient lags. Testing for serial correlation is an important topic in its own right, and so we devote Section 9.4 to it.

[6]The SC penalizes additional lags more heavily than does the AIC and hence leads to a more parsimonious model. It is generally preferred to the AIC that can select a model with too many lags even when the sample size is infinitely large. For details, see Russell Davidson and James McKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, p.676–677.

EXAMPLE 9.8 | Choosing Lag Lengths in an ARDL(p, q) Unemployment Equation

Our objective is to use the SC to select the number of lags for U and the number of lags for G in the equation

$$U_t = \delta + \theta_1 U_{t-1} + \cdots + \theta_p U_{t-p} + \delta_1 G_{t-1} + \cdots + \delta_q G_{t-q} + e_t$$

When computing the SC for a number of possible lag lengths, it is important that the same number of observations is used to estimate each model; otherwise, the sum-of-squared-errors component in the SC will not be comparable across models. Since lagging variables leads to a loss of observations, and the number of observations lost depends on the lag length, care must be exercised when selecting the period for estimation. We consider a maximum of eight lags for both U and G and, to ensure comparability, our estimation period is from 1950Q1 to 2016Q1 for *all models*. Up to eight observations are used for the lags on the right-hand side of each equation, and the first sample value for U_t is always 1950Q1, giving a total of 265 observations. The SC values for $p = 1, 2, 4, 6, 8$ and $q = 0, 1, 2, \ldots, 8$ are displayed in Table 9.5.[7] There are p lags of U and q lags of G. The SC values for $p = 3, 5, 7$ were omitted because they were dominated by those for the other values of p and did not convey any extra information. Because the SC values are negative, the minimizing values for p and q are those that lead to the "largest negative" entry, namely $p = 2$ and $q = 0$, suggesting that the ARDL(2, 0) model $U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + e_t$ is suitable. Other things to notice are that the relatively large increases in the SC if U_{t-2} is dropped and that more than two lags of U_t are not favored by the SC irrespective of the value of q.

Since we have also used an ARDL(2, 1) model with G_{t-1} included, we ask whether there is any evidence to support its inclusion. It turns out that, if we go back and start the sample from 1948Q3, dropping the first two observations to accommodate two lags, the SC values for the ARDL(2, 0) and ARDL(2, 1) models are −2.393 and −2.395, respectively. In this case, there is a slight preference for including G_{t-1}. Moreover, as we have seen from equation (9.42), the coefficient of G_{t-1} is significantly different from zero at a 5% significance level. Its p-value for a zero null hypothesis is 0.014.

TABLE 9.5 SC Values for ARDL(p, q) Unemployment Equation

Lag	SC				
q/p	1	2	4	6	8
0	−1.880	−2.414	−2.391	−2.365	−2.331
1	−2.078	−2.408	−2.382	−2.357	−2.323
2	−2.063	−2.390	−2.361	−2.337	−2.302
3	−2.078	−2.407	−2.365	−2.340	−2.306
4	−2.104	−2.403	−2.362	−2.331	−2.297
5	−2.132	−2.392	−2.353	−2.346	−2.312
6	−2.111	−2.385	−2.346	−2.330	−2.292
7	−2.092	−2.364	−2.325	−2.309	−2.271
8	−2.109	−2.368	−2.327	−2.307	−2.269

9.3.4 Testing for Granger Causality

Granger causality[8] refers to the ability of lags of one variable to contribute to the forecast of another variable. In the context of the ARDL model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t$$

we say that x does not "Granger cause" y if

$$E\left(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-p}, x_{t-1}, x_{t-2}, \ldots x_{t-q}\right) = E\left(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-p}\right)$$

Thus, testing for Granger causality is equivalent to testing

$$H_0 : \delta_1 = 0, \ \ \delta_2 = 0, \ldots, \delta_q = 0$$
$$H_1 : \text{at least one } \delta_i \neq 0$$

[7] The AIC and SC values that are reported are computed using the formulas given in equations (6.43) and (6.44). Your software may provide different values that are based on more general formulas that use a likelihood function. To get the likelihood-based values, you need to add $[1 + \ln(2\pi)] \cong 2.8379$ to the entries in Table 9.4. Adding or subtracting a constant does not change the lag length that minimizes AIC or SC.

[8] Granger, C.W.J. (1969), "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica* 37, 424–38.

It can be performed using the F-test introduced in Chapter 6 for testing joint linear hypotheses. Rejection of H_0 implies x Granger causes y. Note that if x Granger causes y, it does not necessarily imply a direct causal relationship between x and y. It means that having information on past x values will improve the forecast for y. Any causal effect can be an indirect one.

EXAMPLE 9.9 | Does the Growth Rate Granger Cause Unemployment?

To answer this question, we first return to the ARDL(2, 1) model whose estimates were given in equation (9.42). Specifically,

$$\begin{array}{lllll} \hat{U}_t = & 0.3616 + & 1.5331U_{t-1} & -0.5818U_{t-2} & -0.04824G_{t-1} \\ \text{(se)} & (0.0723) & (0.0556) & (0.0556) & (0.01949) \end{array}$$

In this model, testing whether G Granger causes U is equivalent to testing the significance of the coefficient of G_{t-1}. It can be carried out with a t- or an F-test. For example, the F-value is

$$F = t^2 = (0.04824/0.01949)^2 = 6.126$$

It exceeds the 5% critical value of $F_{(0.95,\,1,\,267)} = 3.877$, leading us to conclude that G Granger causes U.

To illustrate how the test works when more than one lag is being tested, consider the following model with four lags of G

$$U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + \delta_1 G_{t-1} + \delta_2 G_{t-2} + \delta_3 G_{t-3} + \delta_4 G_{t-4} + e_t$$

In this model, testing whether G Granger causes U is equivalent to testing

$$H_0: \delta_1 = 0,\ \ \delta_2 = 0,\ \ \delta_3 = 0,\ \ \delta_4 = 0$$
$$H_1: \text{at least one } \delta_i \neq 0$$

The restricted model obtained by assuming that H_0 is true is $U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + e_t$. If we compute an F-value using the restricted and unrestricted sums of squared errors, it is important to make sure that both models use the same number of observations, in this case, 269, for the sample period 1949Q1 to 2016Q1. The F-value for the test is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} = \frac{(23.2471 - 21.3020)/4}{21.3020/(269-7)} = 5.981$$

Because $F = 5.981$ is greater than the 5% critical value $F_{(0.95,\,4,\,262)} = 2.406$, we reject H_0 and conclude that G does Granger cause U.

9.4 Testing for Serially Correlated Errors

Consider again the ARDL(p, q) model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t$$

with $I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\}$ defined as the information set at time $t-1$ and representing all past observations at time t. To keep the notation and exposition relatively simple, suppose $p = q = 1$. One implication of forecasting assumption F2, that all relevant lags have been included in the conditional expectation $E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1}$, is that the errors e_t are serially uncorrelated. For the absence of serial correlation, we require the conditional covariance between any two different errors to be zero. That is, $E(e_t e_s|\mathbf{z}_t, \mathbf{z}_s) = 0$ for all $t \neq s$ where $\mathbf{z}_t = (1, y_{t-1}, x_{t-1})$ denotes all right-hand-side variables in the ARDL model at time t. If $E(e_t e_s|\mathbf{z}_t, \mathbf{z}_s) \neq 0$, then $E(e_t|I_{t-1}) \neq 0$ that, in turn, implies $E(y_t|I_{t-1}) \neq \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1}$. Thus, one way of assessing whether sufficient lags have been included to get the best forecast is to test for serially correlated errors.

Not using the best model for forecasting is not the only implication of serially correlated errors. If $E(e_t e_s|\mathbf{z}_t, \mathbf{z}_s) \neq 0$ for $t \neq s$, then the usual least squares standard errors are invalid. The possibility of invalid standard errors is relevant not just for forecasting equations but also for equations used for policy analysis to be discussed in Section 9.5. For these reasons, testing for serially correlated errors is routine practice when estimating time series regressions. We discuss three tests for this purpose – checking the correlogram of the least squares residuals, a Lagrange multiplier test, and the Durbin–Watson test.

9.4.1 Checking the Correlogram of the Least Squares Residuals

In Section 9.1.2, we saw how the correlogram can be used to examine the nature of the autocorrelations of a time series and to test whether these autocorrelations are significantly different from zero. The autocorrelations for the unemployment and growth series were investigated in Examples 9.2 and 9.3, respectively. In a similar way, we can use the correlogram of the least squares residuals to check for serially correlated errors. Because the errors e_t are unobserved, their correlogram cannot be checked directly. However, we can obtain the least squares residuals $\hat{e}_t = y_t - \hat{\delta} - \hat{\theta}_1 y_{t-1} - \hat{\delta}_1 x_{t-1}$ as estimates of the e_t and examine their autocorrelations. Noting that the mean of the least squares residuals is zero and adapting equation (9.20), we can write the kth order autocorrelation for the residuals as

$$r_k = \frac{\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^{T} \hat{e}_t^2} \tag{9.45}$$

Ideally, for the correlogram to suggest no serial correlation, we like to have $|r_k| < 2/\sqrt{T}$ for $k = 1, 2, \ldots$, the 2 being used to approximate 1.96, the critical value for a 5% significance level. However, occasional significant (but small) autocorrelations at long lags do not constitute strong evidence of autocorrelation and are regarded as acceptable.

EXAMPLE 9.10 | Checking the Residual Correlogram for the ARDL(2, 1) Unemployment Equation

For a first example, we return to the ARDL(2,1) model in (9.42), estimated with 271 observations:

$$\begin{aligned} \hat{U}_t &= 0.3616 + 1.5331 U_{t-1} - 0.5818 U_{t-2} - 0.04824 G_{t-1} \\ (\text{se}) & \quad (0.0723) \quad (0.0556) \qquad (0.0556) \qquad (0.01949) \end{aligned}$$

The autocorrelations for its residuals given in the correlogram in Figure 9.7 are generally small and insignificant. There are exceptions at lags 7, 8, and 17, where the autocorrelations exceed the significance bounds. These correlations are at long lags, barely significant, and relatively small $(r_7 = 0.146, r_8 = -0.130, r_{17} = 0.133)$. It is reasonable to conclude that there is no strong evidence of serial correlation.

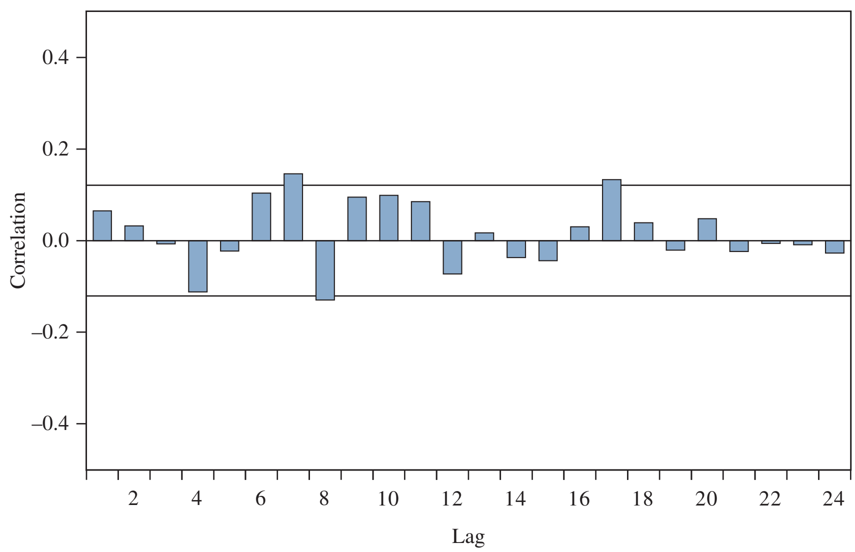

FIGURE 9.7 **Correlogram for residuals from ARDL(2, 1) model.**

EXAMPLE 9.11 | Checking the Residual Correlogram for an ARDL(1, 1) Unemployment Equation

To contrast the outcome in Example 9.10 with one where serial correlation is clearly present, we reestimate the model with U_{t-2} omitted and using 272 observations. If U_{t-2} is an important contributor to the forecasting equation, its omission is likely to lead to serial correlation in the errors. The reestimated equation is

$$\hat{U}_t = 0.4849 + 0.9628U_{t-1} - 0.1672G_{t-1}$$
$$\text{(se)} \quad (0.0842) \quad (0.0128) \qquad (0.0187) \qquad\qquad (9.46)$$

and its correlogram is displayed in Figure 9.8. In this case, the first three autocorrelations are significant, and the first two are moderately large $(r_1 = 0.449, r_2 = 0.313)$. We conclude that the errors are serially correlated. More lags are needed to improve the forecasting specification, and the least squares standard errors given in (9.46) are invalid.

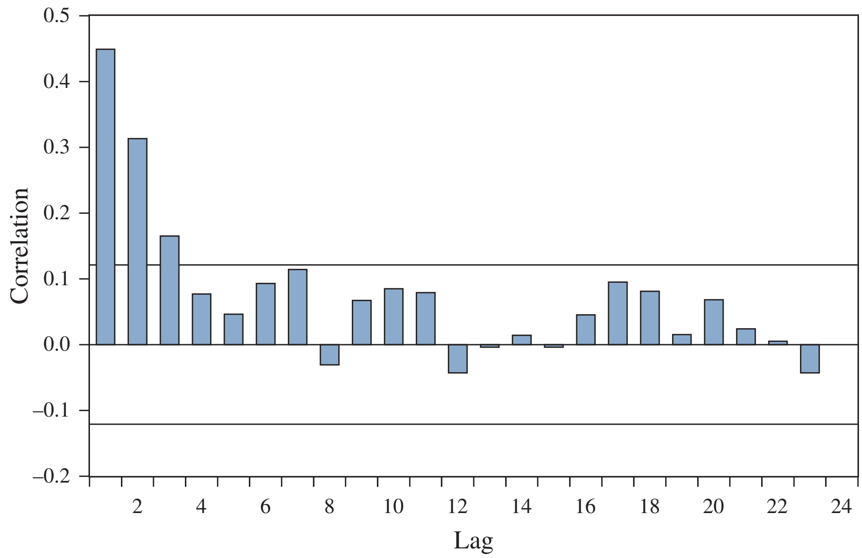

FIGURE 9.8 **Correlogram for residuals from ARDL(1, 1) model.**

9.4.2 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[9] multiplier (LM) tests. In more advanced courses, you will learn the origin of the term LM. Another example was given in Chapter 8 for testing for heteroskedasticity. 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, consider the ARDL(1, 1) model $y_t = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1} + e_t$. The null hypothesis for the test is that the errors e_t are uncorrelated. To express this null hypothesis in terms of restrictions on one or more parameters, we can introduce a model for an alternative hypothesis, with that model describing the possible nature of any autocorrelation. We will consider a number of alternative models.

[9]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.

Testing for AR(1) Errors In the first instance, we consider an alternative hypothesis that the errors are correlated through the AR(1) process $e_t = \rho e_{t-1} + v_t$ where the new errors v_t satisfy the uncorrelated assumption $\text{cov}(v_t, v_s|\mathbf{z}_t, \mathbf{z}_s) = 0$ for $t \neq s$. In the context of the ARDL(1, 1) model, $\mathbf{z}_t = (1, y_{t-1}, x_{t-1})$. Substituting for e_t in the original equation yields

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

Now, if $\rho = 0$, then $e_t = v_t$ and since v_t is not serially correlated, e_t will not be serially correlated. Thus, a test for serial correlation can be set up in terms of the hypotheses $H_0 : \rho = 0$ and $H_1 : \rho \neq 0$. The obvious way to perform this test if e_{t-1} was observable is to regress y_t on y_{t-1}, $x_{t-1,}$ 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, applied econometricians have managed to do it in at least four different ways! One of the variations centers around the treatment of the first observation. To appreciate the issue, suppose that we have 100 observations with which to estimate an ARDL(1, 1) model. Because y and x are both lagged once, an effective sample of 99 observations will be used for estimation. There will be 99 residuals $\hat{e}_t$. Replacing e_{t-1} with $\hat{e}_{t-1}$ in (9.47) means that a further observation will be lost leaving 98 for the test equation. An alternative to losing this last observation is to set the initial value of $\hat{e}_{t-1}$ equal to zero so that 99 observations are retained. Doing so is justified because, when H_0 is true, $E(e_{t-1}|\mathbf{z}_{t-1}) = 0$. This is the approach adopted in the automatic commands of the popular software packages Stata and EViews.

The second variation requires a bit more work. 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 (9.47) as

$$y_t = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1} + \rho \hat{e}_{t-1} + v_t \tag{9.48}$$

Noting that $y_t = \hat{y}_t + \hat{e}_t = \hat{\delta} + \hat{\theta}_1 y_{t-1} + \hat{\delta}_1 x_{t-1} + \hat{e}_t$, we can rewrite (9.48) as

$$\hat{\delta} + \hat{\theta}_1 y_{t-1} + \hat{\delta}_1 x_{t-1} + \hat{e}_t = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1} + \rho \hat{e}_{t-1} + v_t$$

Rearranging this equation yields

$$\begin{aligned}\hat{e}_t &= \left(\delta - \hat{\delta}\right) + \left(\theta_1 - \hat{\theta}_1\right) y_{t-1} + \left(\delta_1 - \hat{\delta}_1\right) x_{t-1} + \rho \hat{e}_{t-1} + v_t \\ &= \gamma_1 + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + \rho \hat{e}_{t-1} + v_t \end{aligned} \tag{9.49}$$

where $\gamma_1 = \delta - \hat{\delta}$, $\gamma_2 = \theta_1 - \hat{\theta}_1$, and $\gamma_3 = \delta_1 - \hat{\delta}_1$. When testing for autocorrelation by testing the significance of the coefficient of $\hat{e}_{t-1}$, one can estimate (9.48) or (9.49). 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 coefficients of y_{t-1} and x_{t-1} will be different, however. In (9.49), we are estimating $\left(\delta - \hat{\delta}\right)$, $\left(\theta_1 - \hat{\theta}_1\right)$, and $\left(\delta_1 - \hat{\delta}_1\right)$, instead of δ, θ_1, and δ_1. The auxiliary regression from which the $T \times R^2$ version of the LM test is obtained is (9.49). Because $\left(\delta - \hat{\delta}\right)$, $\left(\theta_1 - \hat{\theta}_1\right)$, and $\left(\delta_1 - \hat{\delta}_1\right)$ are centered around zero, if (9.49) is a regression with significant explanatory power, that power will come from $\hat{e}_{t-1}$.

If $H_0 : \rho = 0$ is true, then $\text{LM} = T \times R^2$ 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 (9.49). Once again, there are two alternatives depending on whether the first observation is discarded, or $\hat{e}_0$ is set equal to zero.

Testing for MA(1) Errors There are several kinds of models that can be used to try to capture the characteristics of observed sample autocorrelations. These models can be applied to observed time series such as unemployment and the growth rate of GDP or to unobserved errors in a time-series regression model. Up to now only autoregressive models have been discussed. Another useful class of models is what is known as **moving-average** models. You will study these and other models in more depth if you take a time-series course. In Exercise 9.5, you are asked to compare the autocorrelations of an AR(1) model with those of a moving-average model of order one, MA(1). Our task at the moment is to work out a test statistic when an alternative hypothesis of autocorrelation is modeled using the MA(1) process

$$e_t = \phi v_{t-1} + v_t \tag{9.50}$$

The v_t are assumed to be uncorrelated: $\text{cov}\left(v_t, v_s | \mathbf{z}_t, \mathbf{z}_s\right) = 0$ for $t \neq s$. Following the strategy we adopted for the AR(1) error model, combining (9.50) with an ARDL(1,1) model yields

$$y_t = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1} + \phi v_{t-1} + v_t \tag{9.51}$$

Notice that $\phi = 0$ implies $e_t = v_t$, and so we can test for autocorrelation through the hypotheses $H_0: \phi = 0$ and $H_1: \phi \neq 0$. Comparing (9.51) with (9.47), we can see that the test for an MA(1) alternative will be exactly the same as the test for an AR(1) alternative providing we can find an estimate $\hat{v}_{t-1}$. Fortunately, we can use the least squares residual $\hat{e}_{t-1}$ to estimate v_{t-1}, just as we did before. That is, $\hat{v}_{t-1} = \hat{e}_{t-1}$. The reason we can do this is that, when H_0 is true, both errors are the same: $e_t = v_t$. Thus, the test for testing against the alternative of MA(1) errors is identical to the test for an alternative of AR(1) errors. The downside of this result is that, when H_0 is rejected, the LM test does not identify which error model is more suitable.

Testing for Higher Order AR or MA Errors The LM test and its variations can be readily extended to alternative hypotheses that are expressed in terms of higher order AR or MA models. For example, suppose that the model for an alternative hypothesis is either an AR(4) or an MA(4) process. Then

$$\begin{aligned} \text{AR}(4): \quad & e_t = \psi_1 e_{t-1} + \psi_2 e_{t-2} + \psi_3 e_{t-3} + \psi_4 e_{t-4} + v_t \\ \text{MA}(4): \quad & e_t = \phi_1 v_{t-1} + \phi_2 v_{t-2} + \phi_3 v_{t-3} + \phi_4 v_{t-4} + v_t \end{aligned}$$

The corresponding null and alternative hypotheses for each case are

$$\text{AR}(4) \quad \begin{cases} H_0: \psi_1 = 0, \quad \psi_2 = 0, \quad \psi_3 = 0, \quad \psi_4 = 0 \\ H_1: \text{at least one } \psi_i \text{ is nonzero} \end{cases}$$

$$\text{MA}(4) \begin{cases} H_0: \phi_1 = 0, \quad \phi_2 = 0, \quad \phi_3 = 0, \quad \phi_4 = 0 \\ H_1: \text{at least one } \phi_i \text{ is nonzero} \end{cases}$$

The two alternative test equations corresponding to (9.48) and (9.49) are

$$y_t = \delta + \theta_1 y_{t-1} + \delta_1 x_{t-1} + \psi_1 \hat{e}_{t-1} + \psi_2 \hat{e}_{t-2} + \psi_3 \hat{e}_{t-3} + \psi_4 \hat{e}_{t-4} + v_t \tag{9.52}$$

$$\hat{e}_t = \gamma_1 + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + \psi_1 \hat{e}_{t-1} + \psi_2 \hat{e}_{t-2} + \psi_3 \hat{e}_{t-3} + \psi_4 \hat{e}_{t-4} + v_t \tag{9.53}$$

We have used the coefficient notation ψ_i from the AR model, but since the test is the same for both AR and MA alternatives, we could equally well have used ϕ_i from the MA model. One can use an F-test to jointly test the significance of the ψ_i in (9.52) or (9.53), or, use the $\text{LM} = T \times R^2$ test computed from (9.53). When H_0 is true, the latter has a $\chi^2_{(4)}$-distribution. Once again, the initial observations can be dropped or set to zero; there will be a slight difference in results from these two alternatives.

EXAMPLE 9.12 | LM Test for Serial Correlation in the ARDL Unemployment Equation

To illustrate the LM test, we apply the $\chi^2 = T \times R^2$ version of the test to the ARDL unemployment equation. Two models are chosen: the ARDL(1, 1) model whose residual correlogram strongly suggested the existence of serially correlated errors and the ARDL(2, 1) model whose correlogram revealed a few small significant correlations, but otherwise was free from serial correlation. Initial values for the $\hat{e}_t$ lost from lagging were set to zero. Table 9.6 contains the test results for AR(k) or MA(k) alternatives for $k = 1, 2, 3, 4$. There is strong evidence that the errors in the ARDL(1, 1) model are serially correlated. With p-values less than 0.0001, the test soundly rejects a null hypothesis of no serial correlation at all four lags. With the ARDL(2,1) model, the results are not so clear cut. At a 5% significance level, a null hypothesis of no serial correlation is not rejected for alternatives with one lag or four lags, but it is rejected for alternatives with two or three lags. Adding a second lag of U_t to the ARDL(1, 1) model has eliminated a large degree of serial correlation in the errors, but some may still remain. In Exercise 9.19, you are invited to test for serial correlation in the errors after adding more lags of U_t and G_t.

TABLE 9.6 LM Test Results for Serial Correlation in the Errors of the Unemployment Equation

Values of k for AR(k) or MA(k) Alternative	ARDL(1, 1)		ARDL(2, 1)	
	Test value	p-Value	Test Value	p-Value
1	66.90	0.0000	2.489	0.1146
2	73.38	0.0000	6.088	0.0476
3	73.38	0.0000	9.253	0.0261
4	73.55	0.0000	9.930	0.0521

9.4.3 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 AR(1) 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. In addition, unlike the LM and correlogram tests, its distribution no longer holds when the equation contains a lagged dependent variable. A quick rule of thumb, useful when checking your computer output, is that a Durbin–Watson statistic value near 2.0 is compatible with the hypothesis of no serial correlation. Details are provided in Appendix 9A.

9.5 Time-Series Regressions for Policy Analysis

In Section 9.3, we focused on specification, estimation, and use of time-series regressions for forecasting. The main concern was how to use an estimate of an AR conditional expectation

$$E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p}$$

or an ARDL conditional expectation

$$E(y_t|I_{t-1}) = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q}$$

to forecast the future values y_{T+1}, $y_{T+2}, \ldots$, given the information available at the end of the sample period, I_T. In the AR model, the information set was $I_T = \{y_T, y_{T-1}, y_{T-2}, \ldots\}$; for the ARDL model it was $I_T = \{y_T, x_T, y_{T-1}, x_{T-1}, y_{T-2}, x_{T-2}, \ldots\}$. We were not concerned with the interpretation of individual coefficients, and, providing an adequate number of lags of y (or y

and x) was included in the relevant conditional expectation, we were not concerned with omitted variables. Valid forecasts could be obtained from either of the models or one that contains other explanatory variables and their lags. Moreover, because we were using past data to forecast the future, a current value of x was not included in the ARDL model.

Models for policy analysis differ in a number of ways. The individual coefficients are of interest because they might have a causal interpretation, telling us how much the average outcome of a dependent variable changes when an explanatory variable and its lags change. For example, central banks who set interest rates are concerned with how a change in the interest rate will affect inflation, unemployment, and GDP growth, now and in the future. Because we are interested in the current effect of a change, as well as future effects, the current value of explanatory variables can appear in distributed lag or ARDL models. In addition, omitted variables can be a problem if they are correlated with the included variables because then the coefficients may not reflect causal effects.

Interpreting a coefficient as the change in a dependent variable *caused* by a change in an explanatory variable is in line with the emphasis in Chapters 2–8. With the exception of Section 6.3.1, where we discussed the difference between predictive and causal models, and some sections devoted to prediction, the focus in those chapters was on estimating $\beta_k = \partial E(y_t|\mathbf{x}_t)/\partial x_{tk}$ in the model

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + e_t$$

and on how the interpretation of the β_k changes if one or more variables is expressed in terms of logarithms or if there is some other nonlinear relationship between y_t and x_{tk}. Results from these earlier chapters on the estimation of causal effects also hold for time series regressions providing some critical assumptions hold. Under assumptions MR1–MR5 described in Chapter 5, least squares estimates of the β_k are best linear unbiased. However, there are two of these assumptions that can be very restrictive when working within a time-series framework. Recalling that $\mathbf{X}$ is used to denote all observations in all time periods for the right-hand-side variables, those two assumptions are strict exogeneity, $E(e_t|\mathbf{X}) = 0$, and the absence of serial correlation in the errors, $\text{cov}(e_t, e_s|\mathbf{X}) = 0$ for $t \neq s$. Strict exogeneity implies that there are no lagged dependent variables on the right-hand side, ruling out ARDL models. It also means that the errors are uncorrelated with future x values, an assumption that would be violated if x was a policy variable, such as the interest rate, whose setting was influenced by past values of y, such as the inflation rate. The absence of serial correlation implies that variables omitted from the equation, and whose effect is felt through the error term, must not be serially correlated. Given that time series variables are typically autocorrelated, it is likely to be difficult to satisfy this assumption.

The strict exogeneity assumption can be relaxed if we are content to live with large sample properties. In Section 5.7.3, we noted that the assumptions $E(e_t) = 0$ and $\text{cov}(e_t, x_{tk}) = 0$ for all t and k were sufficient for the least squares estimator to be consistent. Thus, we can still proceed if the errors and right-hand-side variables are contemporaneously uncorrelated, an implication of the lesser assumption of **contemporaneous exogeneity**. In the general framework of an ARDL model, the contemporaneous exogeneity assumption can be written as $E(e_t|\mathbf{z}_t) = 0$ where $\mathbf{z}_t$ denotes all right-hand-side variables that could include both lagged x's and lagged y's. Feedback from current and past y to future x is possible under this assumption, and lagged values of y can be included on the right-hand side. However, as we will discover, for both proper interpretation of coefficients and consistency of estimation, we have to be careful about including the correct number of lags and about the context in which lagged values of y and x arise in the equation. Stronger assumptions often have to be made. In Section 9.1.1, we noted that lagged values of y can arise not just in ARDL models but also in transformations of other models: in a model with an AR(1) error and in an IDL model. The special features of these models are considered in Sections 9.5.3 and 9.5.4. For the OLS standard errors to be valid for large sample inference, the serially uncorrelated error assumption $\text{cov}(e_t, e_s|\mathbf{X}) = 0$ for $t \neq s$ can be weakened to $\text{cov}(e_t, e_s|\mathbf{z}_t, \mathbf{z}_s) = 0$ for $t \neq s$, but we do still need to query whether this assumption is realistic in a time-series setting.

In the following four sections, we are concerned with three main issues that add to our time-series regression results from earlier chapters.

1. Interpretation of coefficients of lagged variables in finite and infinite distributed lag models.
2. Estimation and inference for coefficients when the errors are autocorrelated.
3. The assumptions necessary for interpretation and estimation.

To simplify the discussion, we work with models with only one x and its lags, like those specified at the beginning of this chapter in Table 9.1. Our results and conclusions carry over to models with more than one x and their lags.

9.5.1 Finite Distributed Lags

The finite distributed lag model where we are interested in the impact of current and past values of a variable x on current and future values of a variable y can be written as

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \tag{9.54}$$

It is called a *finite* distributed lag because the impact of x on y cuts off after q lags. It is called a *distributed* lag because the impact of a change in x is distributed over future time periods. For the coefficients β_k to represent causal effects, the error term must not be correlated with any omitted variables that are correlated with $\mathbf{x}_t = (x_t, x_{t-1}, \ldots, x_{t-q})$. In particular, since x_t is likely to be autocorrelated, we require e_t not to be correlated with the current and *all past* values of x. This requirement holds if

$$E(e_t | x_t, x_{t-1}, \ldots) = 0 \tag{9.54}$$

It then follows that

$$\begin{aligned} E(y_t | x_t, x_{t-1}, \ldots) &= \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} \\ &= E(y_t | x_t, x_{t-1}, \ldots, x_{t-q}) = E(y_t | \mathbf{x}_t) \end{aligned} \tag{9.55}$$

Once q lags of x have been included in the equation, further lags of x will not have an impact on y.

Given this assumption, a lag-coefficient β_s can be interpreted as the change in $E(y_t | \mathbf{x}_t)$ when x_{t-s} changes by 1 unit, but x is held constant in other periods. Alternatively, if we look forward instead of backward, β_s gives the change in $E(y_{t+s} | \mathbf{x}_t)$ when x_t changes by 1 unit, but x is held constant in other periods. In terms of derivatives

$$\frac{\partial E(y_t | \mathbf{x}_t)}{\partial x_{t-s}} = \frac{\partial E(y_{t+s} | \mathbf{x}_t)}{\partial x_t} = \beta_s \tag{9.56}$$

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 1 unit and then returned to its original level in the next and subsequent periods. Then, using (9.54) 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 1-unit change in x_t is **distributed** over the current and next q periods, from which we get the term distributed lag model. 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 1 unit and then maintained at its new level in subsequent periods $(t + 1), (t + 2), \cdots$. 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** or **cumulative multipliers**. For example, the 2-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 $\sum_{s=0}^{q} \beta_s$.

EXAMPLE 9.13 | Okun's Law

To illustrate the various distributed lag concepts, we introduce an economic model known as Okun's Law.[10] In this model, we again consider a relationship between unemployment and growth of the economy, but we formulate the model differently and use a different data set. Moreover, our purpose is not to forecast unemployment but to investigate the lagged responses of unemployment to growth in the economy. In the basic model for Okun's Law, the change in the unemployment rate from one period to the next depends on the rate of growth of output in the economy:

$$U_t - U_{t-1} = -\gamma(G_t - G_N) \tag{9.57}$$

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.

To write (9.57) in the more familiar notation of the multiple regression model, we denote the change in unemployment by $DU_t = \Delta U_t = U_t - U_{t-1}$, we set $\beta_0 = -\gamma$ and $\alpha = \gamma G_N$, and include an error term

$$DU_t = \alpha + \beta_0 G_t + e_t \tag{9.58}$$

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 (9.58) to include lags of G_t

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \cdots + \beta_q G_{t-q} + e_t \tag{9.59}$$

To estimate this relationship, we use quarterly Australian data on unemployment and the percentage change in gross domestic product (GDP) from quarter 2, 1978 to quarter 2, 2016. These data are stored in the file *okun5_aus*. The time series for DU and G are graphed in Figure 9.9(a) and (b). There are noticeable jumps in the unemployment rate around 1983, 1992, and 2009; they correspond roughly to periods when there was negative growth but with a lag. At this time, we also note that the series appear to be stationary; tools for more rigorous assessment of stationarity are deferred until Chapter 12.

Least squares estimates of the coefficients and related statistics for equation (9.59) are reported in Table 9.7 for lag

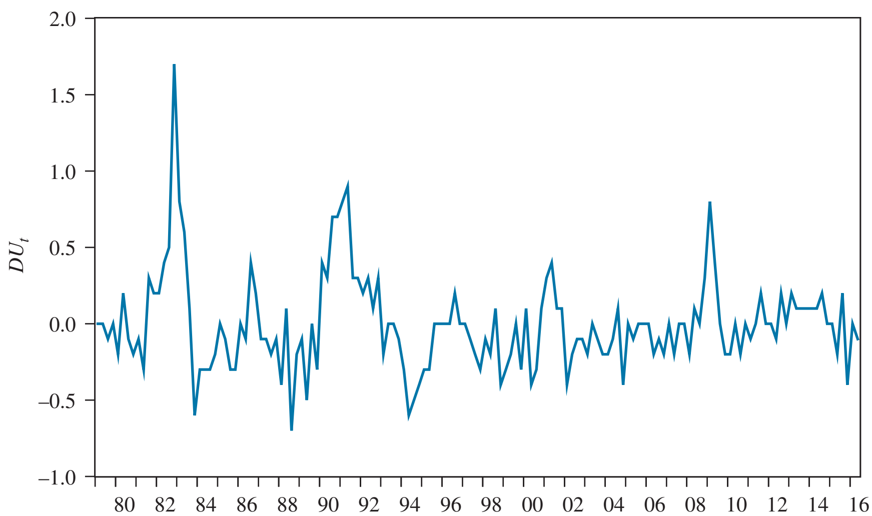

FIGURE 9.9a **Time series for the change in the Australian unemployment rate: 1978Q2 to 2016Q2.**

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

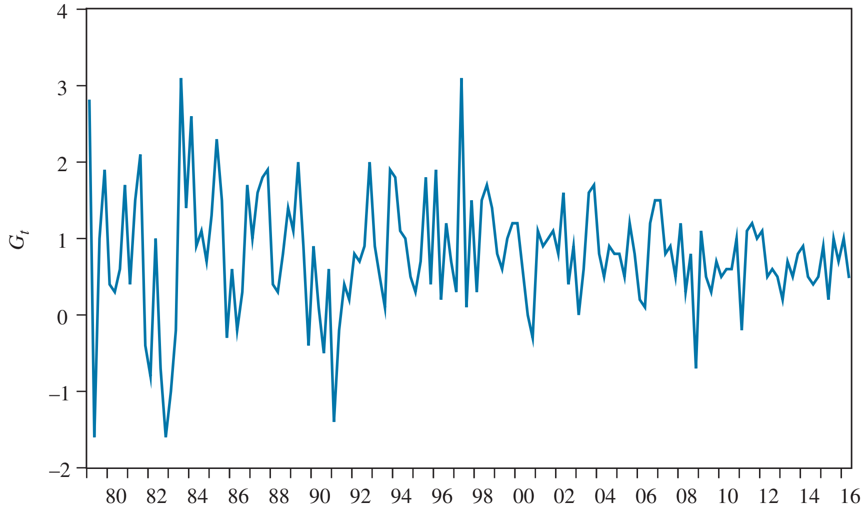

FIGURE 9.9b **Time series for Australian GDP growth: 1978Q2 to 2016Q2.**

lengths $q = 4$ and $q = 5$. 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-5} when $q = 5$. Given the coefficient of this lag is positive and insignificant, we drop G_{t-5} and settle on a model of order $q = 4$ where all coefficients have the expected negative signs and are significantly different from zero.

What do the estimates for lag length 4 tell us? A 1% increase in the growth rate leads to a fall in the expected unemployment rate of 0.13% in the current quarter, a fall of 0.17% in the next quarter and falls of 0.09%, 0.07%, and 0.06% in two, three, and four quarters from now, respectively. These changes represent the values of the impact multiplier and the one- to four-quarter delay multipliers. The interim multipliers, which give the effect of a sustained increase in the growth rate of 1%, are −0.30 for 1 quarter, −0.40 for 2 quarters, −0.47 for 3 quarters, and −0.53 for 4 quarters. Since we have a lag length of four, −0.53 is also the total multiplier. A summary of these values is presented in Table 9.8. Knowledge of them is important for a government that wishes to keep unemployment below a certain level by influencing the growth rate. If we view γ in equation (9.57) as the total effect of a change in output growth, then its estimate is $\hat{\gamma} = -\sum_{s=0}^{4} b_s = 0.5276$. An estimate of the normal growth rate that is needed to maintain a constant unemployment rate is $\hat{G}_N = \hat{\alpha}/\hat{\gamma} = 0.4100/0.5276 = 0.78\%$ per quarter.

TABLE 9.7 **Estimates for Okun's Law Finite Distributed Lag Model**

Lag Length $q = 5$				
Variable	Coefficient	Standard Error	t-Value	p-Value
C	0.3930	0.0449	8.746	0.0000
G_t	−0.1287	0.0256	−5.037	0.0000
G_{t-1}	−0.1721	0.0249	−6.912	0.0000
G_{t-2}	−0.0932	0.0241	−3.865	0.0002
G_{t-3}	−0.0726	0.0241	−3.012	0.0031
G_{t-4}	−0.0636	0.0241	−2.644	0.0091
G_{t-5}	0.0232	0.0240	0.966	0.3355
Observations = 148	$R^2 = 0.503$	$\hat{\sigma} = 0.2258$		
Lag Length $q = 4$				
Variable	Coefficient	Standard Error	t-Value	p-Value
C	0.4100	0.0415	9.867	0.0000
G_t	−0.1310	0.0244	−5.369	0.0000
G_{t-1}	−0.1715	0.0240	−7.161	0.0000
G_{t-2}	−0.0940	0.0240	−3.912	0.0001
G_{t-3}	−0.0700	0.0239	−2.929	0.0041
G_{t-4}	−0.0611	0.0238	−2.563	0.0114
Observations = 149	$R^2 = 0.499$	$\hat{\sigma} = 0.2251$		

TABLE 9.8 **Multipliers for Okun's Law**

Delay Multipliers		Interim Multipliers	
b_0	−0.1310		
b_1	−0.1715	$\sum_{s=0}^{1} b_s$	−0.3025
b_2	−0.0940	$\sum_{s=0}^{2} b_s$	−0.3965
b_3	−0.0700	$\sum_{s=0}^{3} b_s$	−0.4665
b_4	−0.0611	$\sum_{s=0}^{4} b_s$	−0.5276
Total multiplier		$\sum_{s=0}^{4} b_s = -0.5276$	

Assumptions for Finite Distributed Lag Model Before examining some complications that frequently arise with the finite distributed lag model, it is useful to summarize the assumptions that are necessary for OLS estimates to have desirable large sample properties, and the implications of violations of these assumptions. We can also look ahead to what remedies are available to overcome particular violations of assumptions, and their requirements.

FDL1: The time series y and x are stationary and weakly dependent.

FDL2: The finite distributed lag model describing how y responds to current and past values of x can be written as

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t \tag{9.60}$$

FDL3: The error term is exogenous with respect to the current and all past values of x

$$E(e_t | x_t, x_{t-1}, x_{t-2}, \ldots) = 0$$

This assumption ensures

$$E(y_t | x_t, x_{t-1}, x_{t-2}, \ldots) = E(y_t | \mathbf{x}_t)$$

where $\mathbf{x}_t = (x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-q})$. In other words, all relevant lags of x are included in the model. It also implies that there are no omitted variables that are correlated with $\mathbf{x}_t$ and also impact on y_t. This implication raises questions about the Okun's Law example. There are likely to be excluded macro variables that are correlated with GDP growth and that may also impact on the unemployment rate: wage growth, inflation, and interest rates are all possibilities. In the interest of maintaining a relatively simple example, we abstract from these relationships.

FDL4: The error term is not autocorrelated, $\text{cov}(e_t, e_s | \mathbf{x}_t, \mathbf{x}_s) = E(e_t e_s | \mathbf{x}_t, \mathbf{x}_s) = 0$ for $t \neq s$.

FDL5: The error term is homoskedastic, $\text{var}(e_t | \mathbf{x}_t) = E(e_t^2 | \mathbf{x}_t) = \sigma^2$.

Assumptions FDL4 and FDL5 are needed for OLS standard errors, hypothesis tests, and interval estimates to be valid. Since having autocorrelated errors is highly likely, and heteroskedastcity is a possibility, we need to ask how we would proceed when FDL4 and FDL5 are violated. In Chapter 8 when we were faced with the problem of heteroskedastic errors, we considered two possible solutions: (1) using heteroskedasticity consistent robust standard errors for the OLS estimator with no assumptions about the form of the heteroskedasticity being made or (2) making an assumption about the skedastic function and employing a more efficient **generalized least squares** estimator whose standard errors will be valid if the assumption is true. Comparable solutions exist for time series data when FDL4 and FDL5 are violated. It is possible to use the OLS estimator and standard errors known as **HAC (heteroskedasticity and autocorrelation consistent) standard errors**, or **Newey–West standard errors**. Or, we can make some assumption about the nature of the autocorrelation and employ a more efficient generalized squares estimator. In what follows we consider both options. Although the generalized least squares estimator is more efficient, it comes with a cost. In addition to having to make an assumption about the form of the autocorrelation, an exogeneity assumption that is stricter than FDL3 must be made, whereas for OLS with **HAC standard errors**, FDL3 is sufficient.

9.5.2 HAC Standard Errors

To explain the nature of heteroskedasticity and autocorrelation consistent standard errors within a simplified framework, we drop the lagged x's from (9.60), and consider the simple regression model

$$y_t = \alpha + \beta_0 x_t + e_t$$

From Appendix 8A, the least squares estimator for β_0 can be written as

$$b_0 = \beta_0 + \sum_{t=1}^{T} w_t e_t = \beta_0 + \frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\, e_t}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2} = \beta_0 + \frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\, e_t}{s_x^2} \tag{9.61}$$

where s_x^2 is the sample variance for x, using T as the divisor. When e_t was homoskedastic and uncorrelated, we used this result to show that the variance of b_0, conditional on all observations $\mathbf{X}$, is given by (see equation (2.15))

$$\text{var}\left(b_0 | \mathbf{X}\right) = \frac{\sigma_e^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{\sigma_e^2}{Ts_x^2}$$

For a result that was not conditional on $\mathbf{X}$, we obtained the large sample approximate variance for b_0 from the variance of its asymptotic distribution. This variance is given by $\text{var}\left(b_0\right) = \sigma_e^2 / T\sigma_x^2$ and uses the fact that s_x^2 is a consistent estimator for σ_x^2. Other terminology is that σ_x^2 is the probability limit of s_x^2, $s_x^2 \xrightarrow{p} \sigma_x^2$ (see Section 5.7, and in particular the discussions following equations (5.34) and (5.35)).

We are now interested in the unconditional variance of b_0 when e_t is both heteroskedastic and autocorrelated. This is a much harder problem. Following similar steps to those sketched out in Section 5.7, we can replace s_x^2 in (9.61) by its probability limit σ_x^2, and $\bar{x}$ by its probability limit μ_x, and then write the large sample variance of b_0 as

$$\begin{aligned}
\text{var}\left(b_0\right) &= \text{var}\left(\frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \mu_x)\, e_t}{\sigma_x^2}\right) = \frac{1}{T^2\left(\sigma_x^2\right)^2}\text{var}\left(\sum_{t=1}^{T} q_t\right) \\
&= \frac{1}{T^2\left(\sigma_x^2\right)^2}\left[\sum_{t=1}^{T}\text{var}\left(q_t\right) + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\text{cov}\left(q_t, q_{t+s}\right)\right] \\
&= \frac{\sum_{t=1}^{T}\text{var}\left(q_t\right)}{T^2\left(\sigma_x^2\right)^2}\left[1 + \frac{2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\text{cov}\left(q_t, q_{t+s}\right)}{\sum_{t=1}^{T}\text{var}\left(q_t\right)}\right]
\end{aligned} \tag{9.62}$$

where $q_t = \left(x_t - \mu_x\right)e_t$. HAC standard errors are obtained by considering estimators for the quantity outside the big brackets and the quantity inside the big brackets. For the quantity outside the brackets, first note that q_t has a zero mean. Then, using $(T-K)^{-1}\sum_{t=1}^{T}\hat{q}_t^2 = (T-K)^{-1}\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2\hat{e}_t^2$ as an estimator for $\text{var}\left(q_t\right)$, where $\hat{e}_t$ is a least squares residual, $K = 2$ because it is a simple regression, and s_x^2 as an estimator for σ_x^2, an estimator for $\sum_{t=1}^{T}\text{var}\left(q_t\right)/T^2\left(\sigma_x^2\right)^2$ is given by (see Exercise 9.6)

$$\widehat{\text{var}}_{\text{HCE}}\left(b_0\right) = \frac{T\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2\hat{e}_t^2}{(T-K)\left(\sum_{t=1}^{T}\left(x_t - \bar{x}\right)^2\right)^2}$$

Go back and compare this equation with equation (8.9) in Chapter 8. The notation is a little different and the equations are arranged in different ways, but otherwise, they are identical. The quantity outside the brackets in the last line of (9.62) is the large sample unconditional variance of b_0 when there is heteroskedasticity but no autocorrelation. The square root of its estimator $\widehat{\text{var}}_{\text{HCE}}\left(b_0\right)$ is the heteroskedasticity consistent, robust standard error. To get a variance estimator for least squares that is consistent in the presence of both heteroskedasticity and autocorrelation, we need to multiply $\widehat{\text{var}}_{\text{HCE}}\left(b_0\right)$ by an estimator of the quantity in brackets in (9.62). We will denote this quantity as g.

Several estimators for g have been suggested. To discuss the framework in which they are developed, we simplify g as follows (see Exercise 9.6):

$$g = 1 + \frac{2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\text{cov}(q_t, q_{t+s})}{\sum_{t=1}^{T}\text{var}(q_t)} = 1 + \frac{2\sum_{s=1}^{T-1}(T-s)\text{cov}(q_t, q_{t+s})}{T\text{var}(q_t)}$$

$$= 1 + 2\sum_{s=1}^{T-1}\left(\frac{T-s}{T}\right)\tau_s \tag{9.63}$$

where $\tau_s = \text{corr}(q_t, q_{t+s}) = \text{cov}(q_t, q_{t+s})/\text{var}(q_t)$. When there is no serial correlation in the errors, the q_t will also not be autocorrelated, $\tau_s = 0$ for all s, and $g = 1$.To obtain a consistent estimator for g in the presence of autocorrelated errors, the summation in (9.63) is truncated at a lag much smaller than T, the autocorrelations τ_s up to the truncation point are estimated, and the autocorrelations for lags beyond the truncation point are taken as zero. For example, if five autocorrelations are used, the corresponding estimator is

$$\hat{g} = 1 + 2\sum_{s=1}^{5}\left(\frac{6-s}{6}\right)\hat{\tau}_s$$

Alternative estimators differ depending on the number of lags for which the τ_s are estimated and on whether the weights placed on these correlations at each lag are equal to, for example, $(6-s)/6$, or some other alternative. Because there are a large number of possibilities, you will discover that different software packages may yield different HAC standard errors; moreover, 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. Given a suitable estimator $\hat{g}$, the large sample estimator for the variance of b_0, allowing for both heteroscedasticity and autocorrelation in the errors, is

$$\widehat{\text{var}}_{\text{HAC}}(b_0) = \widehat{\text{var}}_{\text{HCE}}(b_0) \times \hat{g}$$

This analysis extends to the finite distributed lag model with q lags and indeed to any time series regression involving stationary variables. The HAC standard errors are given by the square roots of the estimated HAC variances. In Exercise 9.20, you are invited to check whether the errors in the FDL model for Okun's Law in Example 9.13 are autocorrelated and whether using HAC standard errors has an impact on inferences about the multipliers. In Example 9.14 that follows we investigate the impact of serial correlation on the coefficient standard errors for a Phillips curve.

EXAMPLE 9.14 | A Phillips Curve

The Phillips curve has a long history in macroeconomics as a tool for describing the relationship between inflation and unemployment.[11] Our starting point is the model

$$INF_t = INF_t^E - \gamma(U_t - U_{t-1}) \tag{9.64}$$

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 assume that inflationary

[11] For a historical review of the development of different versions, see Gordon, R.J. (2008), "The History of the Phillips Curve: An American Perspective", http://nzae.org.nz/wp-content/uploads/2011/08/nr1217302437.pdf, Keynote Address at the Australasian Meetings of the Econometric Society.

expectations are constant over time and set $\alpha = INF_t^E$. In addition, we set $\beta_0 = -\gamma$, and add an error term, in which case the Phillips curve can be written as the simple regression model

$$INF_t = \alpha + \beta_0 DU_t + e_t \qquad (9.65)$$

The data used for estimating (9.65) are quarterly Australian data from 1987, Quarter 1 to 2016, Quarter 1, a total of 117 observations, stored in the data file *phillips5_aus*. 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 9.10; the time series for the change in the unemployment rate was previously graphed in Figure 9.9(a). Tests for assessing whether these series are stationary are set as exercises in Chapter 12.

The correlogram of the residuals from least squares estimation of (9.65) is presented in Figure 9.11; approximate 5% significance bounds for the autocorrelations are plotted at $\pm 2/\sqrt{117} = \pm 0.185$. There is evidence of moderate correlations at lags 1–5, and smaller ones at lags 6 and 8. To examine the impact of the autocorrelated errors, in Table 9.9, we report the least squares estimates, and conventional (OLS), HCE and HAC standard errors, *t*-values, and

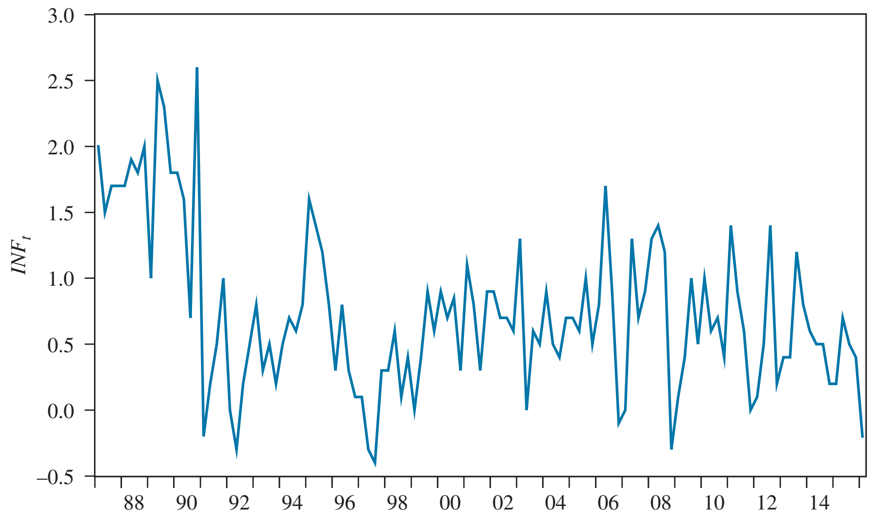

FIGURE 9.10 **Time series for the Australian inflation rate: 1987Q1 to 2016Q1.**

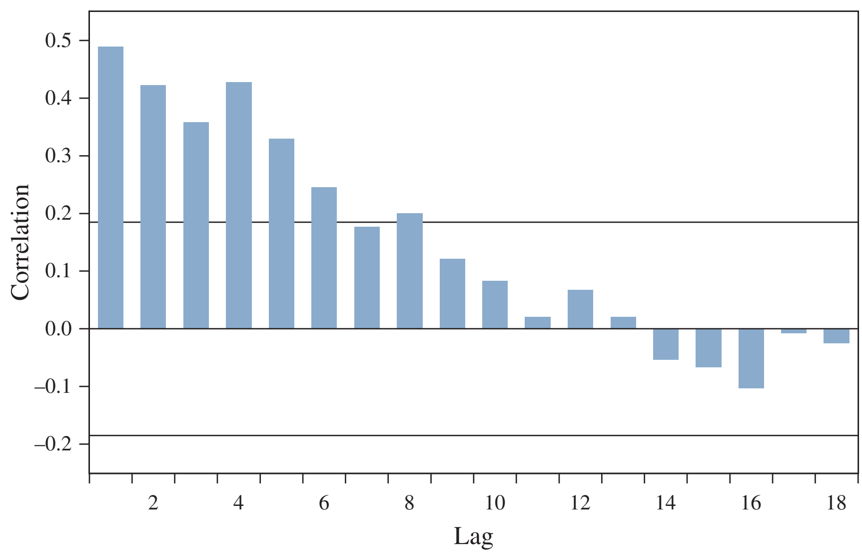

FIGURE 9.11 **Correlogram for least squares residuals from Phillips curve.**

TABLE 9.9 **A Comparison of Conventional (OLS), HCE, and HAC Standard Errors**

Variable	OLS estimate	Standard error			t-value			One-tail p-value		
		OLS	HCE	HAC	OLS	HCE	HAC	OLS	HCE	HAC
C	0.7317	0.0561	0.0569	0.0915	13.05	12.86	7.99	0.0000	0.0000	0.0000
DU	−0.3987	0.2061	0.2632	0.2878	−1.93	−1.51	−1.39	0.0277	0.0663	0.0844

p-values.[12] The HAC standard errors that allow for autocorrelation and heteroskedasticity are larger than the HCE standard errors that allow only for heteroskedacticity, and the HCE standard errors are larger than the conventional OLS ones that allow for neither heteroskedasticity nor autocorrelation. Thus, ignoring the autocorrelation and heteroskedasticity overstates the reliability of the least squares estimates. Overstating their reliability means that interval estimates will be narrower than they should be and we are more likely to reject a true null hypotheses. Using $t_{(0.975,\,115)} = 1.981$, 95% interval estimates for β_0 are $(-0.8070,\ 0.0096)$ with conventional standard errors and $(-0.9688,\ 0.1714)$ with HAC standard errors. With conventional standard errors, a one-tail test and a 5% significance level, we reject $H_0\!:\!\beta_2 = 0$. With HCE or HAC standard errors, we do not reject H_0.

9.5.3 Estimation with AR(1) Errors

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 how to estimate the parameters of the regression model when one such assumption is made, that of AR(1) errors. To keep the exposition free from excessive algebra, we again consider the simple regression model

$$y_t = \alpha + \beta_0 x_t + e_t \tag{9.66}$$

This model can be extended to include extra lags from an FDL model and other variables. The AR(1) error model is given by

$$e_t = \rho e_{t-1} + v_t \qquad |\rho| < 1 \tag{9.67}$$

with the v_t assumed to be uncorrelated random errors with zero mean and constant variances. That is,

$$E(v_t|x_t, x_{t-1}, \ldots) = 0 \qquad \text{var}(v_t|x_t) = \sigma_v^2 \qquad \text{cov}(v_t, v_s|x_t, x_s) = 0 \quad \text{for } t \neq s$$

The assumption $|\rho| < 1$ is required for e_t and y_t to be stationary. From the assumptions about the v_t, we can derive the mean, variance, and autocorrelations for e_t. Conditional on all x's (current, past, and future), it can be shown that e_t has zero mean, constant variance $\sigma_e^2 = \sigma_v^2/(1-\rho^2)$, and autocorrelations $\rho_k = \rho^k$. Thus, the population correlogram that describes the special autocorrelation structure implied by an AR(1) model is $\rho, \rho^2, \rho^3, \ldots$. Because $-1 < \rho < 1$, the AR(1) autocorrelations decline geometrically as the lag increases, eventually becoming negligible. Since there is only one lag of e in the equation $e_t = \rho e_{t-1} + v_t$, you might be surprised to find that autocorrelations at lags greater than one, although declining, are still nonzero.

[12]The HAC standard errors were computed by EViews using a Bartlett kernel, a Newey–West fixed bandwidth of 5, and a degrees of freedom adjustment.

The correlation persists because each e_t depends on all past values of the errors v_t through the equation (see Appendix 9B).[13]

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

Nonlinear Least Squares Estimation

To estimate the AR(1) model described by (9.67) and (9.68), we note, from equation (9.15) in Section 9.1.1, that these equations can be combined and rewritten in the form

$$y_t = \alpha(1 - \rho) + \rho y_{t-1} + \beta_0 x_t - \rho\beta_0 x_{t-1} + v_t \tag{9.68}$$

If you are wondering how we get this equation, go back and check out Section 9.1.1. Why is (9.68) useful for estimation? We have transformed the original model in (9.66) with the autocorrelated error term e_t into a new model given by (9.68) 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 (α, β_0, ρ) that minimize the sum of squares of uncorrelated errors $S_v = \sum_{t=2}^{T} v_t^2$. The least squares estimator that minimizes the sum of squares of the correlated errors $S_e = \sum_{t=1}^{T} e_t^2$ is not minimum variance and its standard errors are not correct. However, minimizing the sum of squares of uncorrelated errors, S_v, yields an estimator that, in large samples, is best and whose standard errors are correct. Note that this result is in line with earlier practice in the book. The least squares estimator used in Chapters 2 through 7 minimizes a sum of squares of uncorrelated errors.

There is, however, an important distinctive feature about the transformed model in (9.68). Note that the coefficient of x_{t-1} is equal to $-\rho\beta_0$, which is the negative product of ρ (the coefficient of y_{t-1}) and β_0 (the coefficient of x_t). This fact means that, although (9.68) 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 (α, β_0, ρ). The usual linear least squares formulas cannot be obtained using calculus to find the values of (α, β_0, ρ) that minimize S_v. Nevertheless, we can still proceed using **nonlinear least squares** to obtain estimates. Nonlinear least squares was introduced in Chapter 6.6. Instead of using formulas to calculate estimates, it uses a numerical procedure to find the estimates that minimize the least squares function.

Generalized Least Squares Estimation

To introduce an alternative estimator for (α, β_0, ρ) in the AR(1) error model, we rewrite (9.68) as

$$y_t - \rho y_{t-1} = \alpha(1 - \rho) + \beta_0 \left(x_t - \rho x_{t-1}\right) + v_t \tag{9.69}$$

Defining $y_t^* = y_t - \rho y_{t-1}$, $\alpha^* = \alpha(1 - \rho)$ and $x_t^* = x_t - \rho x_{t-1}$, (9.69) becomes

$$y_t^* = \alpha^* + \beta_0 x_t^* + v_t \qquad t = 2, 3, \ldots, T \tag{9.70}$$

If ρ was known, values for the transformed variables y_t^* and x_t^* could be calculated, and least squares applied to (9.70) to find estimates $\hat{\alpha}^*$ and $\hat{\beta}_0$. An estimate for the original intercept is $\hat{\alpha} = \hat{\alpha}^*/(1 - \rho)$. This procedure is analogous to that introduced in Section 8.4 where a model with heteroskedastic errors was transformed to one with homoskedastic errors. In that case, the least squares estimator applied to transformed variables y^* and x^* was known as a generalized least squares estimator. Here, we have transformed a model with autocorrelated errors into one with uncorrelated errors. The transformed variables y_t^* and x_t^* are different from those in the heteroscedasticity error case, but, once again, least squares applied to the transformed variables is known as generalized least squares.

Of course, ρ is not known and must be estimated. When the transformed variables are computed using an estimate of ρ, say $\hat{\rho}$, and least squares is applied to these transformed variables, the resulting estimator for α and β_0 is known as a feasible generalized least squares estimator. There are direct parallels with this estimator and the feasible generalized least squares estimator

[13]See Appendix 9B for the derivations.

introduced in Section 8.5. In Section 8.5, parameters in the skedastic function had to be estimated to transform the variables. Here, the parameter in the autocorrelated error model, ρ, needs to be estimated in order to transform the variables.

There are a number of possible estimators for ρ. A simple one is to use r_1 from the sample correlogram. Another one is the least-squares estimate of ρ in a regression of the OLS residuals on their lags. The steps for obtaining the feasible generalized least squares estimator for α and β_0 using this estimator for ρ are as follows:

1. Find least-squares estimates a and b_0 from the equation $y_t = \alpha + \beta_0 x_t + e_t$.
2. Compute the least squares residuals $\hat{e}_t = y_t - a - b_0 x_t$.
3. Estimate ρ by applying least squares to the equation $\hat{e}_t = \rho\hat{e}_{t-1} + \hat{v}_t$. Call this estimate $\hat{\rho}$.
4. Compute values of the transformed variables $y_t^* = y_t - \hat{\rho}y_{t-1}$ and $x_t^* = x_t - \hat{\rho}x_{t-1}$.
5. Apply least squares to the transformed equation $y_t^* = \alpha^* + \beta_0 x_t^* + v_t$.

These steps can also be implemented in an iterative manner. If $\hat{\alpha}$ and $\hat{\beta}_0$ are the estimates obtained in step 5, new residuals can be obtained from $\hat{e}_t = y_t - \hat{\alpha} - \hat{\beta}_0 x_t$, steps 3–5 can be repeated using results from these new residuals, and the process can be continued until the estimates converge. The resulting estimator is often called the **Cochrane–Orcutt** estimator.[14]

Assumptions and Properties

Let's pause and take stock of where we are in Section 9.5. In the finite distributed lag model under assumptions FDL1-FDL5, the least squares estimator is consistent, it is minimum variance in large samples, and the usual OLS t-, F-, and χ^2-tests are valid in large samples. However, time-series data are such that assumptions FDL4 (the errors are not autocorrelated) and FDL5 (homoskedasticity), particularly FDL4, might not hold. When FDL4 and FDL5 are violated, the least squares estimator is still consistent, but its usual variance and covariance estimates and standard errors are not correct, leading to invalid t-, F-, and χ^2-tests. One solution to this problem is to use the HAC estimator for variances and covariances and the corresponding HAC standard errors. The least squares estimator is no longer minimum variance when FDL4 and/or FDL5 do not hold, but using HAC variance and covariance estimates means that t-, F-, and χ^2-tests will be valid. Although we examined the use of HAC standard errors in the context of a simple regression model with no lags, they are equally applicable for a finite distributed lag model that includes lags.

A second solution to violation of FDL4 is to assume a specific model for the autocorrelated errors and to use an estimator that is minimum variance for that model. We showed how the parameters of a simple regression model with AR(1) errors can be estimated by (1) nonlinear least squares or (2) feasible generalized least squares. Under two extra conditions, both of these techniques yield a consistent estimator that is minimum variance in large samples, with valid t-, F-, and χ^2-tests. The first extra condition that is needed to achieve these properties is that the AR(1) error model is suitable for modeling the autocorrelated error. We can, however, guard against a failure of this condition using HAC standard errors following nonlinear least squares or feasible generalized least squares estimation. Doing so will ensure t-, F-, and χ^2-tests are valid despite the wrong choice for an autocorrelated error model. The second extra condition is a stronger exogeneity assumption than that in FDL3. To explore this second requirement, consider estimation of α, β_0, and ρ from the nonlinear least squares equation

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta_0 x_t - \rho\beta_0 x_{t-1} + v_t$$

The exogeneity assumption comparable to FDL3 is

$$E\left(v_t | x_t, x_{t-1}, x_{t-2}, \ldots\right) = 0$$

[14] A modification of this process that includes a transformation of the first observation is called the Prais–Winsten estimator. See Exercise 9.7 for details.

Noting that $v_t = e_t - \rho e_{t-1}$, this condition becomes

$$E(e_t - \rho e_{t-1}|x_t, x_{t-1}, x_{t-2}, \ldots) = E(e_t|x_t, x_{t-1}, x_{t-2}, \ldots) - \rho E(e_{t-1}|x_t, x_{t-1}, x_{t-2}, \ldots) = 0$$

Advancing the subscripts in the second term by one period, we can rewrite this condition as

$$E(e_t|x_t, x_{t-1}, x_{t-2}, \ldots) - \rho E(e_t|x_{t+1}, x_t, x_{t-1}, \ldots) = 0$$

For this equation to be true for all possible values of ρ, we require $E(e_t|x_t, x_{t-1}, x_{t-2}, \ldots) = 0$ and $E(e_t|x_{t+1}, x_t, x_{t-1}, \cdots) = 0$. Now, from the law of iterated expectations, $E(e_t|x_{t+1}, x_t, x_{t-1}, \cdots) = 0$ implies $E(e_t|x_t, x_{t-1}, x_{t-2}, \ldots) = 0$. Thus, the exogeneity requirement necessary for nonlinear least squares to be consistent, and it is the same for feasible generalized least squares, is

$$E(e_t|x_{t+1}, x_t, x_{t-1}, \ldots) = 0 \tag{9.71}$$

This requirement implies that e_t and x_{t+1} cannot be correlated. It rules out instances where x_{t+1} is set by a policymaker (such as a central banker setting an interest rate) in response to an error shock in the previous period. Thus, while modeling the autocorrelated error may appear to be a good strategy in terms of improving the efficiency of estimation, it could be at the expense of consistency if the stronger exogeneity assumption is not met. Using least squares with HAC standard errors does not require this stronger assumption.

Modeling of more general forms of autocorrelated errors with more than one lag requires e_t to be uncorrelated with x values further than one period into the future. A stronger exogeneity assumption that accommodates these more general cases and implies (9.71) is the strict exogeneity assumption $E(e_t|\mathbf{X}) = 0$, where $\mathbf{X}$ includes all current, past and future values of the explanatory variables. For general modeling of autocorrelated errors, we replace FDL3 with this assumption.

EXAMPLE 9.15 | The Phillips Curve with AR(1) Errors

In this example, we obtain estimates of the Phillips curve introduced in Example 9.14 under the assumption that its errors can be modeled with an AR(1) process. The data file is *phillips5_aus*. We can, at the outset, conjecture that an AR(1) model might be inadequate. Returning to the correlogram of the least squares residuals in Figure 9.11, the first four sample autocorrelations are $r_1 = 0.489$, $r_2 = 0.358$, $r_3 = 0.422$, and $r_4 = 0.428$. They do not decline exponentially, nor approximately so. Values that start from $r_1 = 0.489$ and decline in line with the properties of an AR(1) model are $r_2 = 0.489^2 = 0.239$, $r_3 = 0.489^3 = 0.117$, and $r_4 = 0.489^4 = 0.057$. Nevertheless, we illustrate the AR(1) error model with this example and later, in Exercise 9.21, explore how we might improve it. Both the nonlinear least squares (NLS) and feasible generalized least squares (FGLS) estimates are reported in Table 9.10, along with the least squares (OLS) estimates and HAC standard errors reproduced from Table 9.9. The NLS and FGLS estimates and their standard errors are almost identical, and the estimates are also similar to those from OLS. The NLS and FGLS standard errors for estimates of β_0 are smaller than the corresponding OLS HAC standard error, perhaps representing an efficiency gain from modeling the autocorrelation. However, one must be cautious with interpretations like this because standard errors are estimates of standard deviations, not the unknown standard deviations themselves.

TABLE 9.10 Phillips Curve Estimates from AR(1) Error Model

Parameter	OLS		NLS		FGLS	
	Estimate	HAC Standard Error	Estimate	Standard Error	Estimate	Standard Error
α	0.7317	0.0915	0.7028	0.0963	0.7029	0.0956
β_0	−0.3987	0.2878	−0.3830	0.2105	−0.3830	0.2087
ρ			0.5001	0.0809	0.4997	0.0799

9.5.4 Infinite Distributed Lags

The finite distributed lag model introduced in Section 9.5.1 assumed that the effect of changes in an explanatory variable x on a dependent variable y cuts off after a finite number of lags q. One way of avoiding the need to specify a value for q is to consider an IDL model where y depends on lags of x that go back into the indefinite past, namely,

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \cdots + e_t \tag{9.72}$$

We introduced this model in Section 9.1.1. For it to be feasible, the β_s coefficients must eventually (but not necessarily immediately) decline in magnitude, becoming negligible at long lags. They have the same multiplier interpretations as in the finite distributed lag case. Specifically,

$$\beta_s = \frac{\partial E(y_t|x_t, x_{t-1}, \ldots)}{\partial x_{t-s}} = s \text{ period delay multiplier}$$

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

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

For the total multiplier, we assume the infinite sum converges to a finite value.

Geometrically Declining Lags An obvious disadvantage of the IDL model is its infinite number of parameters. To estimate the lag coefficients in (9.72) with a finite sample of data, some kind of restrictions need to be placed on those coefficients. In Section 9.1.1, we showed that insisting the coefficients decline geometrically through the restrictions $\beta_s = \lambda^s \beta_0$, for $0 < \lambda < 1$, led to the ARDL(1, 0) equation

$$y_t = \delta + \theta y_{t-1} + \beta_0 x_t + v_t \tag{9.73}$$

where $\delta = \alpha(1 - \lambda)$, $\theta = \lambda$, and $v_t = e_t - \lambda e_{t-1}$. Go back and reread Section 9.1.1 to see how (9.73) was derived. By imposing the restrictions, we have been able to reduce the infinite number of parameters to just three. The **delay multipliers** can be calculated from the restrictions $\beta_s = \lambda^s \beta_0$. Using results on the sum of a geometric progression, the interim multipliers are given by

$$\sum_{j=0}^{s} \beta_j = \beta_0 + \beta_0\lambda + \beta_0\lambda^2 + \cdots + \beta_0\lambda^s = \frac{\beta_0(1 - \lambda^{s+1})}{1 - \lambda}$$

and the total multiplier is given by

$$\sum_{j=0}^{\infty} \beta_j = \beta_0 + \beta_0\lambda + \beta_0\lambda^2 + \cdots = \frac{\beta_0}{1 - \lambda}$$

Estimating (9.73) poses some difficulties. If we assume that the original errors e_t are not autocorrelated, then $v_t = e_t - \lambda e_{t-1}$ will be correlated with y_{t-1}, which means $E(v_t|y_{t-1}, x_t) \neq 0$; the least squares estimator will be inconsistent. To see that v_t and y_{t-1} are correlated, note that they both depend on e_{t-1}. It is clear that $v_t = e_t - \lambda e_{t-1}$ depends on e_{t-1}. To see that y_{t-1} also depends on e_{t-1}, we lag (9.72) by one period,

$$y_{t-1} = \alpha + \beta_0 x_{t-1} + \beta_1 x_{t-2} + \beta_2 x_{t-3} + \beta_3 x_{t-4} + \cdots + e_{t-1}$$

Assuming, as we have done in the past, that $E(e_t|x_t, x_{t-1}, x_{t-2}, \ldots) = 0$, meaning we cannot predict e_t given current and past values of x, we have

$$\begin{aligned} E(v_t y_{t-1}|x_{t-1}, x_{t-2}, \ldots) &= E\left[(e_t - \lambda e_{t-1})(\alpha + \beta_0 x_{t-1} + \beta_1 x_{t-2} + \cdots + e_{t-1})|x_{t-1}, x_{t-2}, \ldots\right] \\ &= E\left[(e_t - \lambda e_{t-1})\, e_{t-1}|x_{t-1}, x_{t-2}, \ldots\right] \end{aligned}$$

$$= E(e_t e_{t-1} | x_{t-1}, x_{t-2}, \ldots) - \lambda E(e_{t-1}^2 | x_{t-1}, x_{t-2}, \ldots)$$
$$= -\lambda \text{var}(e_{t-1} | x_{t-1}, x_{t-2}, \ldots)$$

where we have used $E(e_t e_{t-1} | x_{t-1}, x_{t-2}, \ldots) = 0$ from the assumption that e_t and e_{t-1} are conditionally uncorrelated.

One possible consistent estimator for (9.73) is the instrumental variable estimator to be discussed in Chapter 10. It turns out that x_{t-1} is a suitable instrument for y_{t-1}. You are encouraged to think of this as an example when you get to Chapter 10.

There is one special case where least squares applied to (9.73) is a consistent estimator. The inconsistency problem arises because the v_t follow the autocorrelated MA(1) process $v_t = e_t - \lambda e_{t-1}$ **and** y_{t-1} appears on the right side of the equation. The v_t are no longer autocorrelated if the e_t follow the AR(1) process $e_t = \lambda e_{t-1} + u_t$, with the **same** parameter λ, and with the u_t being uncorrelated. In this case, we have

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

Since u_t is not autocorrelated, it will not be correlated with y_{t-1}, and so correlation between y_{t-1} and the error is no longer a source of inconsistency for least squares estimation. Clearly, there is a need to check whether $e_t = \lambda e_{t-1} + u_t$ is a reasonable assumption. A test for this purpose has been proposed by McClain and Wooldridge.[15] Details follow.

Testing for Consistency in the ARDL Representation of an IDL Model

The development of this test starts from the assumption that the errors e_t in the IDL model follow an AR(1) process $e_t = \rho e_{t-1} + u_t$ with parameter ρ that can be different from λ and tests the hypothesis $H_0 : \rho = \lambda$. Under the assumption that ρ and λ are different

$$v_t = e_t - \lambda e_{t-1} = \rho e_{t-1} + u_t - \lambda e_{t-1} = (\rho - \lambda) e_{t-1} + u_t$$

Then, equation (9.73) becomes

$$y_t = \delta + \lambda y_{t-1} + \beta_0 x_t + (\rho - \lambda) e_{t-1} + u_t \tag{9.74}$$

The test is based on whether or not an estimate of the error e_{t-1} adds explanatory power to the regression.

The steps are as follows:

1. Compute the least squares residuals from (9.74) under the assumption that H_0 holds

$$\hat{u}_t = y_t - \left(\hat{\delta} + \hat{\lambda} y_{t-1} + \hat{\beta}_0 x_t\right), \quad t = 2, 3, \ldots, T$$

2. Using the least squares estimate $\hat{\lambda}$ from step 1, and starting with $\hat{e}_1 = 0$, compute recursively $\hat{e}_t = \hat{\lambda}\hat{e}_{t-1} + \hat{u}_t$, $t = 2, 3, \ldots, T$.
3. Find the R^2 from a least squares regression of $\hat{u}_t$ on y_{t-1}, x_t and $\hat{e}_{t-1}$.
4. When H_0 is true, and assuming that u_t is homoskedastic, $(T-1) \times R^2$ has a $\chi^2_{(1)}$ distribution in large samples.

Note that $\hat{u}_t$ can be viewed as equal to y_t after y_{t-1} and x_t have been partialled out. Thus, if the regression in step 3 has significant explanatory power, it will come from $\hat{e}_{t-1}$.

We have described this test in the context of a model with geometrically declining lag weights that leads to an ARDL(1, 0) model with only one lag of y. It can also be performed for ARDL(p, q)

[15] McClain, K.T. and J.M. Wooldridge (1995), "A simple test for the consistency of dynamic linear regression in rational distributed lag models," *Economics Letters*, 48, 235–240.

models where $p > 1$. In such instances, the null hypothesis is that the coefficients in an AR(p) error model for e_t are equal to the ARDL coefficients on the lagged y's, extra lags are included in the test procedure, and the chi-square statistic has p degrees of freedom; it is equal to the number of observations used to estimate the test equation multiplied by that equation's R^2.

EXAMPLE 9.16 | A Consumption Function

Suppose that consumption expenditure C is a linear function of "permanent" income Y^*

$$C_t = \omega + \beta Y_t^*$$

Permanent income is unobserved. We will assume that it consists of a trend term and a geometrically weighted average of observed current and past incomes, $Y_t, Y_{t-1}, \ldots$

$$Y_t^* = \gamma_0 + \gamma_1 t + \gamma_2\left(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \lambda^3 Y_{t-3} + \cdots\right)$$

where $t = 0, 1, 2, \ldots$ is the trend term. In this model, consumers anticipate that their income will trend, presumably upwards, adjusted by a weighted average of their past incomes. For reasons that will become apparent in Chapter 12, it is convenient to consider a differenced version of the model where we relate the change in consumption $DC_t = C_t - C_{t-1}$ to the change in actual income $DY_t = Y_t - Y_{t-1}$. This version of the model can be written as

$$\begin{aligned}
DC_t = C_t - C_{t-1} &= \left(\omega + \beta Y_t^*\right) - \left(\omega + \beta Y_{t-1}^*\right) = \beta\left(Y_t^* - Y_{t-1}^*\right) \\
&= \beta\Big\{\gamma_0 + \gamma_1 t + \gamma_2\left(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \lambda^3 Y_{t-3} + \cdots\right) \\
&\quad - \Big[\gamma_0 + \gamma_1(t-1) + \gamma_2\big(Y_{t-1} + \lambda Y_{t-2} + \lambda^2 Y_{t-3} \\
&\quad + \lambda^3 Y_{t-4} + \cdots\big)\Big]\Big\} \\
&= \beta\gamma_1 + \beta\gamma_2\left(DY_t + \lambda DY_{t-1} + \lambda^2 DY_{t-2} + \lambda^3 DY_{t-3} + \cdots\right)
\end{aligned}$$

Setting $\alpha = \beta\gamma_1$ and $\beta_0 = \beta\gamma_2$ and adding an error term, this equation, in more familiar notation, becomes

$$DC_t = \alpha + \beta_0\left(DY_t + \lambda DY_{t-1} + \lambda^2 DY_{t-2} + \lambda^3 DY_{t-3} + \cdots\right) + e_t \tag{9.75}$$

Its ARDL(1, 0) representation is

$$DC_t = \delta + \lambda DC_{t-1} + \beta_0 DY_t + v_t \tag{9.76}$$

To estimate this model, we use quarterly data on Australian consumption expenditure and national disposable income from 1959Q3 to 2016Q3, stored in the data file *cons_inc*. Estimating (9.76) yields

$$\begin{aligned}
\widehat{DC}_t &= 478.6 + 0.3369 DC_{t-1} + 0.0991 DY_t \\
\text{(se)} \quad &\quad (74.2) \quad (0.0599) \qquad\quad (0.0215)
\end{aligned}$$

The delay multipliers from this model are 0.0991, 0.0334, 0.0112, …. The total multiplier is $0.0991/(1 - 0.3369) = 0.149$. At first, these values may seem low for what could be interpreted as a marginal propensity to consume. However, because a trend term is included in the model, we are measuring departures from that trend. The LM test for serial correlation in the errors described in Section 9.4.2 was conducted for lags 1, 2, 3, and 4; in each case, a null hypothesis of no serial correlation was not rejected at a 5% significance level. To see if this lack of serial correlation in the errors could be attributable to an AR(1) model with parameter λ for the errors in (9.75), the steps for the test in the previous subsection were followed, yielding a test value of $\chi^2 = (T-1) \times R^2 = 227 \times 0.00025 = 0.057$. Given the 5% significance level for a $\chi^2_{(1)}$-distribution is 3.84, we fail to reject the null hypothesis that the errors in the IDL representation can be described by the process $e_t = \lambda e_{t-1} + v_t$. Put another way, there is no evidence to suggest that the existence of an MA(1) error of the form $v_t = e_t - \lambda e_{t-1}$ is a source of inconsistency in the estimation of (9.76).

Deriving Multipliers from an ARDL Representation The geometrically declining lag model is a convenient one if we believe the lag weights do in fact satisfy, or approximately satisfy, the restrictions $\beta_s = \lambda^s \beta_0$. However, there are many other lag patterns that may be realistic. The largest impact of a change in an explanatory variable may not be felt immediately; the lag weights may increase at first and then decline. How do we decide what might be reasonable restrictions to impose? Instead of beginning with the IDL representation and choosing restrictions a priori, an alternative strategy is to begin with an ARDL representation whose lags have been chosen using conventional model selection criteria and to derive the restrictions on the IDL model implied by the chosen ARDL model. Specifically, we first estimate the finite number of θ's and δ's from an ARDL model

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

For these estimates to be compatible with the infinite number of β's in the IDL model

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

restrictions have to be placed on the β's. The strategy is to find expressions for the β's in terms of the θ's and δ's such that equations (9.77) and (9.78) are equivalent. One way to do so is to use recursive substitution, substituting out the lagged dependent variables on the right-hand side of (9.77), and going back indefinitely. This process becomes messy very quickly, however, particularly when there are several lags. Our task for the general case is made much easier if we can master some heavy 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

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

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. It will be easier if we continue in terms of a specific example.

EXAMPLE 9.17 | Deriving Multipliers for an Infinite Lag Okun's Law Model

In Example 9.13, using data from the file *okun5_aus*, we estimated a finite distributed lag model for Okun's Law, with the change in unemployment DU_t related to the current value and four lags of GDP growth, $G_t, G_{t-1}, \ldots, G_{t-4}$. Suppose, instead, that we wanted to entertain an IDL with values for G going back into the indefinite past. The estimates in Table 9.7 suggest a geometrically declining lag would be inappropriate. The estimated coefficient for G_{t-1} is larger (in absolute value) than that for G_t and then the coefficients decline. To decide on what might be a suitable lag distribution, we begin by estimating an ARDL model. After experimenting with different values for p and q, taking into consideration significance of the coefficient estimates and the possibility of serial correlation in the errors, we settled on the ARDL(2, 1) model

$$DU_t = \delta + \theta_1 DU_{t-1} + \theta_2 DU_{t-2} + \delta_0 G_t + \delta_1 G_{t-1} + v_t \tag{9.81}$$

Using the lag operator notation in (9.80), this equation can be written as

$$\left(1 - \theta_1 L - \theta_2 L^2\right) DU_t = \delta + \left(\delta_0 + \delta_1 L\right) G_t + v_t \tag{9.82}$$

Now suppose that it is possible to define an inverse of $\left(1 - \theta_1 L - \theta_2 L^2\right)$, that we write as $\left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}$, which is such that

$$\left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}\left(1 - \theta_1 L - \theta_2 L^2\right) = 1$$

This concept is a bit abstract, but we do not have to figure the inverse out. 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 $\left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}$ yields

$$\begin{aligned} DU_t &= \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}\delta \\ &\quad + \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1} \times \left(\delta_0 + \delta_1 L\right) G_t \\ &\quad + \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1} v_t \end{aligned} \tag{9.83}$$

This representation is useful because we can equate it with the IDL 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

$$\alpha = \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}\delta \tag{9.85}$$

$$\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots = \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1}\left(\delta_0 + \delta_1 L\right) \tag{9.86}$$

$$e_t = \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1} v_t \tag{9.87}$$

Equation (9.85) can be used to derive α in terms of θ_1, θ_2, and δ, and equation (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 $\left(1 - \theta_1 L - \theta_2 L^2\right)$ to obtain $\left(1 - \theta_1 L - \theta_2 L^2\right)\alpha = \delta$. Then, recognizing that the lag of a constant is the same constant ($L\alpha = \alpha$), we have

$$\left(1 - \theta_1 - \theta_2\right)\alpha = \delta \quad \text{and} \quad \alpha = \frac{\delta}{1 - \theta_1 - \theta_2}$$

Turning now to the β's, we multiply both sides of (9.86) by $\left(1 - \theta_1 L - \theta_2 L^2\right)$ to obtain

$$\begin{aligned}\delta_0 + \delta_1 L &= \left(1 - \theta_1 L - \theta_2 L^2\right)\left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\right)\\ &= \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\\ &\quad -\theta_1\beta_0 L - \theta_1\beta_1 L^2 - \theta_1\beta_2 L^3 - \cdots\\ &\quad\quad - \theta_2\beta_0 L^2 - \theta_2\beta_1 L^3 - \cdots\\ &= \beta_0 + \left(\beta_1 - \theta_1\beta_0\right)L + \left(\beta_2 - \theta_1\beta_1 - \theta_2\beta_0\right)L^2\\ &\quad + \left(\beta_3 - \theta_1\beta_2 - \theta_2\beta_1\right)L^3 + \cdots \end{aligned} \tag{9.88}$$

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

$$\begin{aligned}\delta_0 + \delta_1 L + 0L^2 + 0L^3 &= \beta_0 + \left(\beta_1 - \theta_1\beta_0\right)L + \left(\beta_2 - \theta_1\beta_1 - \theta_2\beta_0\right)L^2\\ &\quad + \left(\beta_3 - \theta_1\beta_2 - \theta_2\beta_1\right)L^3 + \cdots \end{aligned} \tag{9.89}$$

Equating coefficients of like powers in L yields

$$\begin{aligned}\delta_0 &= \beta_0\\ \delta_1 &= \beta_1 - \theta_1\beta_0\\ 0 &= \beta_2 - \theta_1\beta_1 - \theta_2\beta_0\\ 0 &= \beta_3 - \theta_1\beta_2 - \theta_2\beta_1\end{aligned}$$

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

$$\begin{aligned}\beta_0 &= \delta_0\\ \beta_1 &= \delta_1 + \theta_1\beta_0\\ \beta_j &= \theta_1\beta_{j-1} + \theta_2\beta_{j-2} \quad \text{for} \quad j \geq 2\end{aligned} \tag{9.90}$$

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 equation (9.88) which, in its general form, is

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

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.

EXAMPLE 9.18 | Computing the Multiplier Estimates for the Infinite Lag Okun's Law Model

Using the data file *okun5_aus*, the estimated ARDL(2,1) model for Okun's Law is

$$\begin{aligned}\widehat{DU}_t &= \underset{(0.0328)}{0.1708} + \underset{(0.0767)}{0.2639}DU_{t-1} + \underset{(0.0720)}{0.2072}DU_{t-2}\\ &\quad - \underset{(0.0244)}{0.0904}G_t - \underset{(0.0252)}{0.1296}G_{t-1}\end{aligned} \tag{9.92}$$

(se)

Using the relationships in (9.90), the impact multiplier and the delay multipliers for the first 4 quarters are given by[16]

$$\begin{aligned}\hat{\beta}_0 &= \hat{\delta}_0 = -0.0904\\ \hat{\beta}_1 &= \hat{\delta}_1 + \hat{\theta}_1\hat{\beta}_0 = -0.129647 - 0.263947 \times 0.090400\\ &= -0.1535\\ \hat{\beta}_2 &= \hat{\theta}_1\hat{\beta}_1 + \hat{\theta}_2\hat{\beta}_0 = -0.263947 \times 0.153508\\ &\quad - 0.207237 \times 0.090400 = -0.0593\end{aligned}$$

[16] In the calculations, we carry the values to six decimal places to minimize rounding error.

$$\hat{\beta}_3 = \hat{\theta}_1\hat{\beta}_2 + \hat{\theta}_2\hat{\beta}_1 = -0.263947 \times 0.059252 - 0.207237 \times 0.153508 = -0.0475$$

$$\hat{\beta}_4 = \hat{\theta}_1\hat{\beta}_3 + \hat{\theta}_2\hat{\beta}_2 = -0.263947 \times 0.047452 - 0.207237 \times 0.059252 = -0.0248$$

An increase in GDP growth leads to a fall in unemployment. The effect increases from the current quarter to the next quarter, declines dramatically after that and then gradually 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 9.12 for lags up to 10 quarters.

To estimate the total multiplier that is given by $\sum_{j=0}^{\infty} \beta_j$, we can sum the progressions implied by (9.90), but an easier way is to assume the process is in long-run equilibrium with no changes in DU and G, and to examine the effect of a change in G on the long-run equilibrium. Being in log-run equilibrium means we can ignore the time subscript and the error term in (9.92), giving

$$DU = 0.1708 + 0.2639DU + 0.2072DU - 0.0904G - 0.1296G$$

or

$$DU = \frac{0.1708 - (0.0904 + 0.1296)G}{1 - 0.2639 - 0.2072} = 0.3229 - 0.4160G$$

The total multiplier is given by $d(DU)/dG = -0.416$. The sum of the lag coefficients in Figure 9.12 is $\sum_{s=0}^{10} \hat{\beta}_s = -0.414$; most of the impact of a change in G is felt in the first 10 quarters. An estimate of the normal growth rate that is needed to maintain a constant rate of unemployment is $\hat{G}_N = -\hat{\alpha}/\sum_{j=0}^{\infty} \hat{\beta}_j = 0.3229/0.416 = 0.78\%$. The total multiplier estimate from the finite distributed lag model was higher in absolute value at -0.528, but the estimate of the normal growth rate was the same at 0.78%.

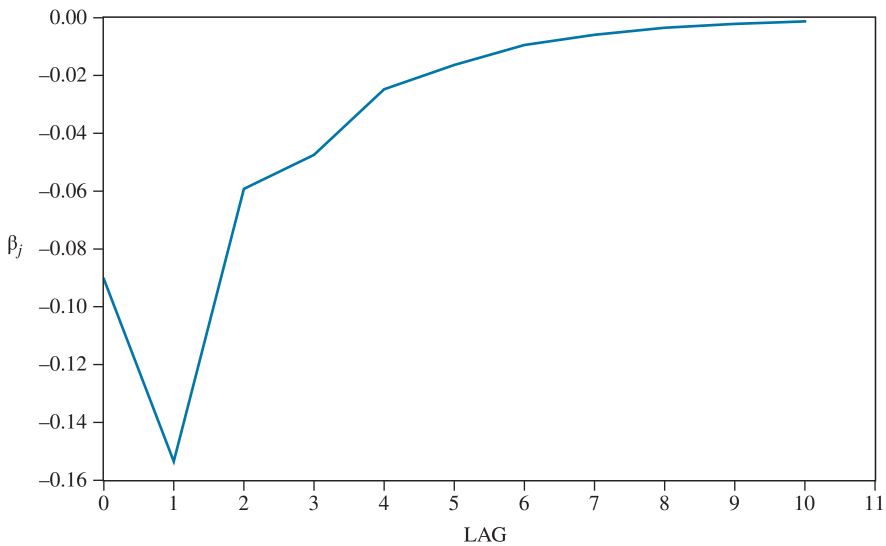

FIGURE 9.12 **Lag distribution from Okun's Law ARDL(2, 1) model.**

The Error Term

In Example 9.18, we used least squares to estimate the ARDL model and conveniently ignored the error term. The question we need to ask is whether the error term will be such that the least squares estimator is consistent. In equation (8.47), we found that

$$e_t = \left(1 - \theta_1 L - \theta_2 L^2\right)^{-1} v_t$$

Multiplying both sides of this equation by $\left(1 - \theta_1 L - \theta_2 L^2\right)$ gives

$$\left(1 - \theta_1 L - \theta_2 L^2\right) e_t = v_t$$
$$e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} = v_t$$
$$e_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + v_t$$

In the general ARDL(p, q) model, this equation becomes

$$e_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_p e_{t-p} + v_t \tag{9.93}$$

For v_t to be uncorrelated, which is required for least squares estimation of the ARDL model to be consistent, the errors e_t must satisfy (9.93). That is, they must follow an AR(p) process with the same coefficients as in the AR component of the ARDL model. The test for consistency of least squares described earlier in the context of the geometric lag model can be extended to the general case.

EXAMPLE 9.19 | Testing for Consistency of Least Squares Estimation of Okun's Law

The starting point for this test is the assumption that the errors e_t in the IDL representation follow an AR(2) process

$$e_t = \psi_1 e_{t-1} + \psi_2 e_{t-2} + v_t$$

with the v_t being uncorrelated. Then, given the ARDL representation

$$DU_t = \delta + \theta_1 DU_{t-1} + \theta_2 DU_{t-2} + \delta_0 G_t + \delta_1 G_{t-1} + v_t \tag{9.94}$$

the null hypothesis is $H_0 : \psi_1 = \theta_1,\ \psi_2 = \theta_2$. To find the test statistic, we compute $\hat{e}_t = \hat{\theta}_1 \hat{e}_{t-1} + \hat{\theta}_2 \hat{e}_{t-2} + \hat{u}_t$ where the $\hat{u}_t$ are the residuals from the estimated equation in (9.92). Then, regressing $\hat{u}_t$ on a constant, DU_{t-1}, DU_{t-2}, G_t, G_{t-1}, $\hat{e}_{t-1,}$ and $\hat{e}_{t-2}$ yields $R^2 = 0.02089$ and a test value $\chi^2 = (T-3) \times R^2 = 150 \times 0.02089 = 3.13$. The 5% critical value is $\chi^2_{(0.95,2)} = 5.99$ implying we fail to reject H_0 at a 5% significance level. There is not sufficient evidence to conclude that serially correlated errors are a source of inconsistency in least squares estimation of (9.94).

Assumptions for the Infinite Distributed Lag Model Several assumptions underlie least squares estimation of the consumption function and Okun's Law examples. Here we summarize those assumptions and discuss implications of variations of them.

IDL1: The time series y and x are stationary and weakly dependent.

IDL2: The infinite distributed lag model describing how y responds to current and past values of x can be written as

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + e_t \tag{9.95}$$

with $\beta_s \to 0$ as $s \to \infty$.

IDL3: Corresponding to (9.95) is an ARDL(p, q) model

$$y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + v_t \tag{9.96}$$

where $v_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_p e_{t-p}$.

IDL4: The errors e_t are strictly exogenous,

$$E(e_t|\mathbf{X}) = 0$$

where $\mathbf{X}$ includes all current, past, and future values of x.

IDL5: The errors e_t follow the AR(p) process

$$e_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_p e_{t-p} + u_t$$

where

i. u_t is exogenous with respect to current and past values of x and past values of y,

$$E(u_t|x_t, x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, \ldots) = 0$$

ii. u_t is homoskedastic, $\text{var}(u_t|x_t) = \sigma_u^2$

Under assumptions IDL2 and IDL3, expressions for the lag weights β_s in terms of the parameters θ's and δ's can be found by equating coefficients of like powers of the lag operator in the product

$$\delta_0 + \delta_1 L + \delta_2 L^2 + \cdots + \delta_q L^q = \left(1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p\right) \times \left(\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3 + \cdots\right) \tag{9.97}$$

The assumption IDL5 is a very special case of an autocorrelated error model for (9.95) and for that reason we described a test of its validity. It is required for least squares estimation of (9.96) to be consistent. Because the exogeneity assumption IDL5(i) includes all past values of y, it is sufficient to ensure v_t will not be autocorrelated; IDL5(ii) is needed for OLS standard errors to be valid. If IDL5 holds and least squares estimates of (9.96) are used to find estimates of the β's through equation (9.97), strict exogeneity for e_t (IDL4) is required for the β's to have a causal interpretation. This requirement is similar to that for nonlinear least squares and generalized least squares estimation of the autocorrelated error model.

An alternative assumption to IDL5 is

IDL5*: The errors e_t are uncorrelated, $\text{cov}\left(e_t, e_s | x_t, x_s\right) = 0$ for $t \neq s$ and homoskedastic, $\text{var}\left(e_t | x_t\right) = \sigma_e^2$.

In this case, the errors $v_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_p e_{t-p}$ follow an MA(p) process, and least squares estimation of (9.96) is inconsistent. The instrumental variables approach studied in Chapter 10 can be used as an alternative.

Finally, we note that both an FDL model with autocorrelated errors and an IDL model can be transformed to ARDL models. Thus, an issue that arises after estimating an ARDL model is whether to interpret it as an FDL model with autocorrelated errors or an IDL model. An attractive way out of this dilemma is to assume an FDL model and use HAC standard errors. In many cases, an IDL model will be well approximated by an FDL, and using HAC standard errors avoids having to make the restrictive strict exogeneity assumption.

9.6 Exercises

9.6.1 Problems

9.1 a. Show that the mean-squared forecast error $E\left[\left(\hat{y}_{T+1} - y_{T+1}\right)^2 \middle| I_T\right]$ for a forecast $\hat{y}_{T+1}$, that depends only on past information I_T, can be written as

$$E\left[\left(\hat{y}_{T+1} - y_{T+1}\right)^2 \middle| I_T\right] = E\left[\left\{\left(\hat{y}_{T+1} - E\left(y_{T+1} | I_T\right)\right) - \left(y_{T+1} - E\left(y_{T+1} | I_T\right)\right)\right\}^2 \middle| I_T\right]$$

b. Show that $E\left[\left(\hat{y}_{T+1} - y_{T+1}\right)^2 \middle| I_T\right]$ is minimized by choosing $\hat{y}_{T+1} = E\left(y_{T+1} | I_T\right)$.

9.2 Consider the AR(1) model $y_t = \delta + \theta y_{t-1} + e_t$ where $|\theta| < 1$, $E\left(e_t | I_{t-1}\right) = 0$ and $\text{var}\left(e_t | I_{t-1}\right) = \sigma^2$. Let $\bar{y}_{-1} = \sum_{t=2}^{T} y_t / (T-1)$ (the average of the observations on y with the first one missing) and $\bar{y}_{-T} = \sum_{t=2}^{T} y_{t-1} / (T-1)$ (the average of the observations on y with the last one missing).

a. Show that the least squares estimator for θ can be written as

$$\hat{\theta} = \theta + \frac{\sum_{t=2}^{T} e_t \left(y_{t-1} - \bar{y}_{-T}\right)}{\sum_{t=2}^{T} \left(y_{t-1} - \bar{y}_{-T}\right)^2}$$

b. Explain why $\hat{\theta}$ is a biased estimator for θ.
c. Explain why $\hat{\theta}$ is a consistent estimator for θ.

9.3 Consider a stationary model that combines the AR(2) model $y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + e_t$ with an AR(1) error model $e_t = \rho e_{t-1} + v_t$ where $E(v_t | I_{t-1}) = 0$. Show that

$$E(y_t | I_{t-1}) = \delta(1-\rho) + (\theta_1 + \rho)\, y_{t-1} + (\theta_2 - \theta_1 \rho) y_{t-2} - \theta_2 \rho y_{t-3}$$

Why will the assumption $E(y_t | I_{t-1}) = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2}$ be violated if the errors are autocorrelated?

9.4 Consider the ARDL(2, 1) model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \delta_1 x_{t-1} + e_t$$

with auxiliary AR(1) model $x_t = \alpha + \phi x_{t-1} + v_t$, where $I_t = \{y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots\}$, $E(e_t | I_{t-1}) = 0$, $E(v_t | I_{t-1}) = 0$, $\text{var}(e_t | I_{t-1}) = \sigma_e^2$, $\text{var}(v_t | I_{t-1}) = \sigma_v^2$, and v_t and e_t are independent. Assume that sample observations are available for $t = 1, 2, \ldots, T$.

a. Show that the best forecasts for periods $T+1$, $T+2$ and $T+3$ are given by

$$\hat{y}_{T+1} = \delta + \theta_1 y_T + \theta_2 y_{T-1} + \delta_1 x_T$$

$$\hat{y}_{T+2} = \delta + \delta_1 \alpha + \theta_1 \hat{y}_{T+1} + \theta_2 y_T + \delta_1 \phi x_T$$

$$\hat{y}_{T+3} = \delta + \delta_1 \alpha + \delta_1 \phi \alpha + \theta_1 \hat{y}_{T+2} + \theta_2 \hat{y}_{T+1} + \delta_1 \phi^2 x_T$$

b. Show that the variances of the forecast errors are given by

$$\sigma_{f1}^2 = E\left((y_{T+1} - \hat{y}_{T+1})^2 \middle| I_T \right) = \sigma_e^2$$

$$\sigma_{f2}^2 = E\left((y_{T+2} - \hat{y}_{T+2})^2 \middle| I_T \right) = (1 + \theta_1^2)\, \sigma_e^2 + \delta_1^2 \sigma_v^2$$

$$\sigma_{f3}^2 = E\left((y_{T+3} - \hat{y}_{T+3})^2 \middle| I_T \right) = \left((\theta_1^2 + \theta_2)^2 + \theta_1^2 + 1 \right) \sigma_e^2 + \delta_1^2 \left((\theta_1 + \phi)^2 + 1 \right) \sigma_v^2$$

9.5 Let e_t denote the error term in a time series regression. We wish to compare the autocorrelations from an AR(1) error model $e_t = \rho e_{t-1} + v_t$ with those from an MA(1) error model $e_t = \phi v_{t-1} + v_t$. In both cases, we assume that $E(v_t v_{t-s}) = 0$ for $s \neq 0$ and $E(v_t^2) = \sigma_v^2$. Let $\rho_s = E(e_t e_{t-s}) / \text{var}(e_t)$ be the s-th order autocorrelation for e_t. Show that,

a. for an AR(1) error model, $\rho_1 = \rho, \rho_2 = \rho^2, \rho_3 = \rho^3, \ldots$

b. for an MA(1) error model, $\rho_1 = \phi / (1 + \phi^2)\,, \rho_2 = 0, \rho_3 = 0, \ldots$

Describe in words the difference between the two autocorrelation structures.

9.6 This question is designed to clarify some of the results used to explain HAC standard errors.

a. Given that $\widehat{\text{var}}(\hat{q}_t) = (T-2)^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2 \hat{e}_t^2$ and $s_x^2 = T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2$, show that

$$\frac{\sum_{t=1}^{T} \widehat{\text{var}}(\hat{q}_t)}{T^2 (s_x^2)^2} = \frac{T \sum_{t=1}^{T} (x_t - \bar{x})^2 \hat{e}_t^2}{(T-2) \left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^2}$$

b. For $T = 4$, write out all the terms in the summations

$$\text{(i)} \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \text{cov}(q_t, q_{t+s}) \quad \text{and} \quad \text{(ii)} \sum_{s=1}^{T-1} (T-s)\, \text{cov}(q_t, q_{t+s})$$

What assumption is necessary for these two summations to be equal?

c. For the simple regression model $y_t = \alpha + \beta_0 x_t + e_t$ with $E(e_t | x_t) = 0$ show that $\text{cov}(e_t, e_s | x_t, x_s) = 0$ for $t \neq s$ implies $\text{cov}(q_t, q_s) = 0$ where $q_t = (x_t - \mu_x)\, e_t$.

9.7 In Section 9.5.3, we described how a generalized least squares (GLS) estimator for α and β_0 in the regression model $y_t = \alpha + \beta_0 x_t + e_t$, with AR(1) errors $e_t = \rho e_{t-1} + v_t$ and known ρ, can be computed by applying OLS to the transformed model $y_t^* = \alpha^* + \beta_0 x_t^* + v_t$ where $y_t^* = y_t - \rho y_{t-1}$, $\alpha^* = \alpha(1-\rho)$ and $x_t^* = x_t - \rho x_{t-1}$. In large samples, the GLS estimator is minimum variance because the v_t are homoskedastic and not autocorrelated. However, x_t^* and y_t^* can only be found for $t = 2, 3, \ldots, T$. One observation is lost through the transformation. To ensure the GLS estimator is minimum variance in small samples, a transformed observation for $t = 1$ has to be included. Let $e_1^* = \sqrt{1 - \rho^2} e_1$.

a. Using results in Appendix 9B, show that $\text{var}(e_1^*) = \sigma_v^2$ and that e_1^* is uncorrelated with v_t, $t = 2, 3, \ldots, T$.

b. Explain why the result in (a) implies OLS applied to the following transformed model will yield a minimum variance estimator

$$y_t^* = \alpha j_t + \beta_0 x_t^* + e_t^*$$

where $y_t^* = y_t - \rho y_{t-1}$, $j_t = 1 - \rho$, $x_t^* = x_t - \rho x_{t-1}$, and $e_t^* = e_t - \rho e_{t-1} = v_t$ for $t = 2, 3, \ldots, T$, and, for $t = 1$, $y_1^* = \sqrt{1-\rho^2}y_1$, $j_1 = \sqrt{1-\rho^2}$, and $x_1^* = \sqrt{1-\rho^2}x_1$. This estimator, particularly when it is used iteratively with an estimate of ρ, is often known as the Prais–Winsten estimator.

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

$$INVGWTH_t = 4 - 0.4FFRATE_t - 0.6FFRATE_{t-1} - 0.3FFRATE_{t-2} - 0.2FFRATE_{t-3}$$

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

b. Suppose that *FFRATE* is raised by one percentage point to 2% 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 that *FFRATE* is raised to 2% 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?

9.9 Using 157 weekly observations on sales revenue (*SALES*) and advertising expenditure (*ADV*) in millions of dollars for a large department store, the following relationship was estimated

$$\widehat{SALES}_t = 18.74 + 1.006ADV_t + 3.926ADV_{t-1} + 2.372ADV_{t-2}$$

a. How many degrees of freedom are there for this estimated model? (Take into account the observations lost through lagged variables.)

b. 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?

c. The estimated covariance matrix of the coefficients is

	C	ADV_t	ADV_{t-1}	ADV_{t-2}
C	0.2927	−0.1545	−0.0511	−0.0999
ADV_t	−0.1545	0.4818	−0.3372	0.0201
ADV_{t-1}	−0.0511	−0.3372	0.7176	−0.3269
ADV_{t-2}	−0.0999	0.0201	−0.3269	0.4713

Using a two-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?

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

9.10 Consider the following time series sample of size $T = 10$ on a random variable y_t whose sample mean is $\bar{y} = 0$.

t	1	2	3	4	5	6	7	8	9	10
y_t	1	4	8	5	4	−3	0	−5	−9	−5

a. Use a hand calculator or spreadsheet to compute the sample autocorrelations

$$r_1 = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_t^2} \qquad r_2 = \frac{\sum_{t=3}^{T} y_t y_{t-2}}{\sum_{t=1}^{T} y_t^2} \qquad r_3 = \frac{\sum_{t=4}^{T} y_t y_{t-3}}{\sum_{t=1}^{T} y_t^2}$$

b. Using a 5% significance level, separately test whether r_1, r_2, and r_3 are significantly different from zero. Sketch the first three bars of the correlogram. Include the significance bounds.

9.11 Using 250 quarterly observations on U.S. GDP growth (G) from 1947Q2 to 2009Q3, we calculate the following quantities.

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

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

a. Compute the first three autocorrelations $(r_1, r_2, \text{and } r_3)$ 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 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).

9.12 Increases in the mortgage interest rate increase the cost of owning a house and lower the demand for houses. In this question, we use three equations to forecast the monthly change in the number of new one-family houses sold in the United States. In the first equation (XR 9.12.1), the monthly change in the number of houses *DHOMES* is regressed against two lags of the monthly change in the 30-year conventional mortgage rate *DIRATE*. In the second equation (XR 9.12.2), *DHOMES* is regressed against two lags of itself, and in the third equation (XR 9.12.3), two lags of both *DHOMES* and *DIRATE* are included as regressors.

$$DHOMES_t = \delta + \delta_1 DIRATE_{t-1} + \delta_2 DIRATE_{t-2} + e_{1,t} \quad \text{(XR 9.12.1)}$$

$$DHOMES_t = \delta + \theta_1 DHOMES_{t-1} + \theta_2 DHOMES_{t-2} + e_{2,t} \quad \text{(XR 9.12.2)}$$

$$DHOMES_t = \delta + \theta_1 DHOMES_{t-1} + \theta_2 DHOMES_{t-2} + \delta_1 DIRATE_{t-1} + \delta_2 DIRATE_{t-2} + e_{3,t} \quad \text{(XR 9.12.3)}$$

The data used are from January, 1992 (1992M1) to September, 2016 (2016M9). The units of measurement are thousands of new houses for *DHOMES* and percentage points for *DIRATE*. After differencing and allowing for two lags, three observations are lost, resulting in a total of 294 observations that were used to produce the least squares estimates in Table 9.11.

TABLE 9.11 **Coefficient Estimates for Equations for Forecasting New Houses**

	XR 9.12.1		XR 9.12.2		XR 9.12.3	
Dependent variable	$DHOMES_t$	$\hat{e}_{1,t}$	$DHOMES_t$	$\hat{e}_{2,t}$	$DHOMES_t$	$\hat{e}_{3,t}$
C	−0.92	−0.03	0.05	0.05	−1.39	0.65
$DHOMES_{t-1}$			−0.32	0.04	−0.37	0.53
$DHOMES_{t-2}$			−0.10	0.16	−0.11	0.14
$DIRATE_{t-1}$	−46.1	−0.31			−45.6	−0.003
$DIRATEt_{-2}$	−13.2	−1.17			−35.3	30.8
$\hat{e}_{1,t-1}$		−0.39		−0.05		−0.54
$\hat{e}_{1,t-2}$		−0.14		−0.17		0.03
SSE	634312	550482	599720	597568	555967	550770

a. Given $DHOMES_{2016M8} = -54$, $DHOMES_{2016M9} = 18$, $DIRATE_{2016M8} = 0.00$, $DIRATE_{2016M9} = 0.02$, and $DIRATE_{2016M10} = -0.01$, use each of the three estimated equations to find 95% forecast intervals for $DHOMES_{2016M10}$ and $DHOMES_{2016M11}$. Comment on the results.

b. Using a 5% significance level, test for autocorrelated errors in each of the equations.

c. Using a 5% significance level, test whether *DIRATE* Granger causes *DHOMES*.

9.13 Consider the infinite lag representation $y_t = \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + e_t$ for the ARDL model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_3 y_{t-3} + \delta_1 x_{t-1} + v_t$$

a. Show that $\alpha = \delta/(1 - \theta_1 - \theta_3)$, $\beta_0 = 0, \beta_1 = \delta_1$, $\beta_2 = \theta_1\beta_1$, $\beta_3 = \theta_1\beta_2$, and $\beta_s = \theta_1\beta_{s-1} + \theta_3\beta_{s-3}$ for $s \geq 4$.

b. Using quarterly data on U.S. inflation (INF), and the change in the unemployment rate (DU) from 1955Q2 to 2016Q1, we estimate the following version of a Phillips curve

$$\underset{(\text{se})}{\widehat{INF}_t} = \underset{(0.049)}{0.094} + \underset{(0.051)}{0.564INF_{t-1}} + \underset{(0.052)}{0.333INF_{t-3}} - \underset{(0.084)}{0.300DU_{t-1}} \qquad SSE = 48.857$$

c. Using the results in part (a), find estimates of the first 12 lag weights in the infinite lag representation of the estimated Phillips curve in part (b). Graph those weights and comment on the graph.

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

e. Let $\hat{e}_t = 0.564\hat{e}_{t-1} + 0.333\hat{e}_{t-3} + \hat{u}_t$ where the $\hat{u}_t$ are the residuals from the equation in part (b), and the initial values $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$ are set equal to zero. The SSE from regressing $\hat{u}_t$ on a constant, $INF_{t-1}, INF_{t-3}DU_{t-1}, \hat{e}_{t-1}$, and $\hat{e}_{t-3}$ is 47.619. Using a 5% significance level, test the hypothesis that the errors in the infinite lag representation follow the AR(3) process $e_t = \theta_1 e_{t-1} + \theta_3 e_{t-3} + v_t$. The number of observations used in this regression and that in part (b) is 241. What are the implications of this test result?

9.14 Inflationary expectations play an important role in wage negotiations between employers and employees. In this exercise, we examine how inflationary expectations of Australian businesses, collected by National Australia Bank surveys, depend on past inflation rates. The data are quarterly and run from 1989Q3 to 2016Q1. The basic model being estimated is

$$EXPN_t = \alpha + \beta_1 INF_{t-1} + e_t$$

where $EXPN_t$ is the expected percentage price increase for 3 months ahead and INF_{t-1} is the inflation rate in the previous 3 months. The left-hand panel of estimates in Table 9.12 contains OLS estimates of α and β_1 with conventional and HAC standard errors. The right-hand panel contains nonlinear least squares estimates and both sets of standard errors assuming the equation errors follow the AR(1) process $e_t = \rho e_{t-1} + v_t$. The first three sample autocorrelations of the residuals are also reported for each of the estimations.

TABLE 9.12 **Estimates for Inflationary Expectations Model**

	OLS Estimates			AR(1) Error Model		
		OLS	HAC		NLS	HAC
	Coefficients	Standard Errors	Standard Errors	Coefficients	Standard Errors	Standard Errors
α	1.437	0.110	0.147	1.637	0.219	0.195
β_1	0.629	0.120	0.188	0.208	0.074	0.086
ρ				0.771	0.063	0.076
	$r_1 = 0.651$			$r_1 = -0.132$		
	$r_2 = 0.466$			$r_2 = 0.099$		
	$r_3 = 0.445$			$r_3 = -0.136$		
	Observations = 106			Observations = 105		

a. What evidence is there of serial correlation in the errors e_t? What is the impact of any serial correlation on interval estimation of β_1?

b. Is there any evidence of remaining serial correlation in the errors v_t after estimating the model with an AR(1) error?

c. What is the impact of the AR(1) error assumption on the estimate for β_1? Suggest a reason for the large difference in magnitude.

d. Show that the AR(1) error model can be written as

$$EXPN_t = \delta + \theta_1 EXPN_{t-1} + \delta_1 INF_{t-1} + \delta_2 INF_{t-2} + v_t$$

where $\delta = \alpha(1-\rho), \theta_1 = \rho, \delta_1 = \beta_1$ and $\delta_2 = -\rho\beta_1$.

e. Estimating the unconstrained version of the model in part (d) via OLS yields

$$\widehat{EXPN}_t = 0.376 + 0.773 EXPN_{t-1} + 0.206 INF_{t-1} - 0.163 INF_{t-2}$$
$$(\text{se}) \qquad (0.121) \quad (0.070) \qquad\qquad (0.091) \qquad\qquad (0.090)$$

Given that $\text{se}\left(\hat{\theta}_1\hat{\delta}_1 + \hat{\delta}_2\right) = 0.1045$, test the hypothesis $H_0\!:\theta_1\delta_1 = -\delta_2$ using a 5% significance level. What is the implication of this test result?

f. Find estimates for the first four lag coefficients of the infinite distributed lag representation of the equation estimated in part (e).

9.15 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

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

9.6.2 Computer Exercises

9.16 Using tha data file *usmacro*, estimate the ARDL(2, 1) model

$$U_t = \delta + \theta_1 U_{t-1} + \theta_2 U_{t-2} + \delta_1 G_{t-1} + e_t$$

Your estimates should agree with the results given in equation (9.42). Use these estimates to verify the forecast results given in Table 9.4.

9.17 Using the data file *usmacro*, estimate the AR(1) model $G_t = \alpha + \phi G_{t-1} + v_t$. From these estimates and those obtained in Exercise 9.16, use the results from Exercise 9.4 to find point and 95% interval forecasts for U_{2016Q2}, U_{2016Q3}, and U_{2016Q4}.

9.18 Consider the ARDL(p, q) equation

$$U_t = \delta + \theta_1 U_{t-1} + \cdots + \theta_p U_{t-p} + \delta_1 G_{t-1} + \cdots + \delta_q G_{t-q} + e_t$$

and the data in the file *usmacro*. For $p = 2$ and $q = 1$, results from the LM test for serially correlated errors were reported in Table 9.6 for AR(k) or MA(k) alternatives with $k = 1, 2, 3, 4$. The $\chi^2 = T \times R^2$ version of the test, with missing initial values for $\hat{e}_t$ set to zero, was used to obtain those results. Considering again the model with $p = 2$ and $q = 1$, compare the results in Table 9.6 with results from the following alternative versions of the LM test.

1. The $\chi^2 = T \times R^2$ version of the test, with missing initial values for $\hat{e}_t$ dropped.
2. The F-test for the joint significance of lags of $\hat{e}_t$, with missing initial values for $\hat{e}_t$ dropped.
3. The F-test for the joint significance of lags of $\hat{e}_t$, with missing initial values for $\hat{e}_t$ set to zero.

9.19 Consider the ARDL(p, q) equation

$$U_t = \delta + \theta_1 U_{t-1} + \cdots + \theta_p U_{t-p} + \delta_1 G_{t-1} + \cdots + \delta_q G_{t-q} + e_t$$

and the data in the file *usmacro*. For $p = 2$ and $q = 1$, results from the LM test for serially correlated errors were reported in Table 9.6 for AR(k) or MA(k) alternatives with $k = 1, 2, 3, 4$. The $\chi^2 = T \times R^2$ version of the test, with missing initial values for $\hat{e}_t$ set to zero, was used to obtain those results.

a. Using the same test statistic and the same AR and MA alternatives, and a 5% significance level, test for serially correlated errors in the two models, ($p = 4, q = 3$) and ($p = 6, q = 5$).

b. Examine the residual correlograms from the two models in part (a). What do they suggest?

9.20 In Example 9.13, the following finite distributed lag model was estimated for Okun's Law using the data file *okun5_aus*.

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + \beta_3 G_{t-3} + \beta_4 G_{t-4} + e_t$$

a. Find the correlogram of the least squares residuals for this model. Is there any evidence of autocorrelation?

b. Test for autocorrelation in the residuals using the $\chi^2 = T \times R^2$ version of the LM test, with missing initial values for $\hat{e}_t$ set to zero, and lags up to 4. Is there any evidence of autocorrelation?

c. Compare 95% interval estimates for the coefficients obtained using conventional OLS standard errors with those obtained using HAC standard errors.

9.21 In Examples 9.14 and 9.15, we considered the Phillips curve

$$INF_t = INF_t^E - \gamma(U_t - U_{t-1}) + e_t = \alpha + \beta_0 DU_t + e_t$$

where inflationary expectations are assumed to be constant, $INF_t^E = \alpha$, and $\beta_0 = -\gamma$. In Example 9.15, we used data in the file *phillips5_aus* to estimate this model assuming the errors follow an AR(1) model $e_t = \rho e_{t-1} + v_t$. Nonlinear least squares estimates of the model were $\hat{\alpha} = 0.7028$, $\hat{\beta}_0 = -0.3830$, and $\hat{\rho} = 0.5001$. The equation from these estimates can be written as the following ARDL representation (see equation (9.68))

$$\begin{aligned}\widehat{INF}_t &= \hat{\alpha}(1 - \hat{\rho}) + \hat{\rho} INF_{t-1} + \hat{\beta}_0 DU_t - \hat{\rho}\hat{\beta}_0 DU_{t-1} \\ &= 0.7028 \times (1 - 0.5001) + 0.5001 INF_{t-1} - 0.3830 DU_t + (0.5001 \times 0.3830)\, DU_{t-1} \\ &= 0.3513 + 0.5001 INF_{t-1} - 0.3830 DU_t + 0.1915 DU_{t-1}\end{aligned} \qquad \text{(XR 9.21.1)}$$

Instead of assuming that this ARDL(1, 1) model is a consequence of an AR(1) error, another possible interpretation is that inflationary expectations depend on actual inflation in the previous quarter, $INF_t^E = \delta + \theta_1 INF_{t-1}$. If DU_{t-1} is retained because of a possible lagged effect, and we change notation so that it is line with what we are using for a general ARDL model, we have the equation

$$INF_t = \delta + \theta_1 INF_{t-1} + \delta_0 DU_t + \delta_1 DU_{t-1} + e_t \qquad \text{(XR 9.21.2)}$$

a. Find least squares estimates of the coefficients in (XR 9.21.2) and compare these values with those in (XR 9.21.1). Use HAC standard errors.

b. Reestimate (XR 9.21.2) after dropping DU_{t-1}. Why is it reasonable to drop DU_{t-1}?

c. Now, suppose that inflationary expectations depend on inflation in the previous quarter and inflation in the same quarter last year, $INF_t^E = \delta + \theta_1 INF_{t-1} + \theta_4 INF_{t-4}$. Estimate the model that corresponds to this assumption.

d. Is there empirical evidence to support the model in part (c)? In your answer, consider (i) the residual correlograms from the equations estimated in parts (b) and (c), and the significance of coefficients in the complete ARDL(4, 0) model that includes INF_{t-2} and INF_{t-3}.

9.22 Using the data file *phillips5_aus*, estimate the equation

$$INF_t = \delta + \theta_1 INF_{t-1} + \theta_4 INF_{t-4} + \delta_0 DU_t + e_t$$

Assuming that the unemployment rate in 2016Q2, 2016Q3 and 2016Q4 remains constant at 6%, use the estimated equation to find 95% forecast intervals for the inflation rate in those quarters.

9.23 Using the data file *phillips5_aus*, estimate the equation

$$INF_t = \delta + \theta_1 INF_{t-1} + \theta_4 INF_{t-4} + \delta_0 DU_t + v_t$$

a. Find the first eight lag weights (delay multipliers) of the infinite distributed lag representation that corresponds to this model. What is the total multiplier?

b. Using a 5% significance level, test the hypothesis that the error term in the infinite distributed lag representation follows the AR(4) process $e_t = \theta_1 e_{t-1} + \theta_4 e_{t-4} + v_t$.

9.24 In Example 9.16, we considered a geometrically declining infinite distributed lag model to describe the relationship between the change in consumption $DC_t = C_t - C_{t-1}$ and the change in income $DY_t = Y_t - Y_{t-1}$. In this exercise, we consider instead a finite distributed lag model of the form

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

a. Use the observations in the data file *cons_inc* to estimate this model assuming $q = 4$. Use HAC standard errors. Comment on (i) the distribution of the lag weights and (ii) the significance of your estimates at a 5% significance level.

b. Reestimate the equation, dropping the lags whose coefficients were not significant in part (a). Use HAC standard errors. Have there been any substantial changes in the estimates and standard errors of the coefficients of the retained lags?

c. Using an LM test with two lags, test for autocorrelation in the errors of the equation estimated in part (b). Is the use of HAC standard errors justified?

d. Assume that the errors follow the AR(1) process $e_t = \rho e_{t-1} + v_t$ with the usual assumptions on v_t. Transform the model estimated in part (b) into one that can be estimated using nonlinear least squares.

e. Use nonlinear least squares to estimate the model derived in part (d). Use HAC standard errors. Compare these estimates and their standard errors with those obtained in part (b).

f. Using the results from part (e), find an estimate for the total multiplier and its standard error. Compare these values with those obtained for the model in Example 9.16. (You will need to estimate the model in Example 9.16 to work out the standard error of its total multiplier.)

9.25 a. Using observations on the change in consumption $DC_t = C_t - C_{t-1}$ and the change in income $DY_t = Y_t - Y_{t-1}$ from 1959Q3 to 2015Q4, obtained from the data file *cons_inc*, estimate the following two models

$$DC_t = \delta + \theta_1 DC_{t-1} + \delta_0 DY_t + e_{1t}$$
$$DC_t = \alpha + \beta_0 DY_t + \beta_3 DY_{t-3} + e_{2t}$$

b. Use each model estimated in part (a) to forecast consumption C in 2016Q1, 2016Q2, and 2016Q3.

c. Use the mean-square criterion $\sum_{t=2016Q1}^{2016Q3}\left(\hat{C}_t - C_t\right)^2$ to compare the out-of-sample predictive ability of the two models.

9.26 Using time series data on five different countries, Atkinson and Leigh[17] examine changes in inequality measured as the percentage income share ($SHARE$) held by those with the top 1% of incomes. A subset of their annual data running from 1921 to 2000 can be found in the data file *inequality*.

a. It is generally recognized that inequality was high prior to the great depression, then declined during the depression and World War II, increasing again toward the end of the sample period. To capture this effect, use the observations on New Zealand to estimate the following model with a quadratic trend

$$SHARE_t = \beta_1 + \beta_2 YEAR_t + \beta_3 YEAR_t^2 + e_t$$

where $YEAR_t$ is defined as $1 = 1921$, $2 = 1922, \ldots, 80 = 2000$. Plot the observations on $SHARE$ and the fitted quadratic trend. Does the trend capture the general direction of the changes in $SHARE$?

b. Find the correlogram of the least-squares residuals from the equation estimated in part (a). How many of the autocorrelations (up to lag 15) are significantly different from zero at a 5% level of significance?

c. Reestimate the equation in (a) using HAC standard errors. How do they compare with the conventional standard errors? Using first the conventional coefficient covariance matrix, and then the HAC covariance matrix, find 95% interval estimates for the expected share in 2001. That is, $E(SHARE|YEAR = 81) = \beta_1 + 81\beta_2 + 81^2\beta_3$. Compare the two intervals.

d. Assuming that the errors in (a) follow the AR(1) error process $e_t = \rho e_{t-1} + v_t$, show that the model can be rewritten as [*Hint*: $YEAR_{t-1} = YEAR_t - 1$]

$$SHARE_t = \beta_1 - \rho\left(\beta_1 - \beta_2 + \beta_3\right) + \rho SHARE_{t-1} + \left[\beta_2 - \rho\left(\beta_2 - 2\beta_3\right)\right]YEAR_t + \beta_3(1-\rho)YEAR_t^2 + v_t$$

e. Estimate the equation in part (d) using nonlinear least squares. Plot the quadratic trend and compare it with that obtained in part (a).

f. Estimate the following equation using OLS and use the estimates of δ_1, δ_2, δ_3, and ρ to retrieve estimates of β_1, β_2, and β_3. How do they compare with the nonlinear least squares estimates obtained in part (e)?

$$SHARE_t = \delta_1 + \rho SHARE_{t-1} + \delta_2 YEAR_t + \delta_3 YEAR_t^2 + v_t$$

[17] Atkinson, A.B. and A. Leigh (2013), "The Distribution of Top Incomes in Five Anglo-Saxon Countries over the Long Run", *Economic Record*, 89, 1–17.

g. Find the correlogram of the least-squares residuals from the equation estimated in part (f). How many of the autocorrelations (up to lag 15) are significantly different from zero at a 5% level of significance?

h. Using the equation estimated in part (f), find a 95% interval estimate for the expected share in 2001. That is, $E\left(SHARE_{2001}|YEAR = 81, SHARE_{2000} = 8.25\right)$. Compare this interval with those obtained in part (c).

9.27 Reconsider the data file *inequality* used in Exercise 9.26 and the model in part (a) of that exercise but include the median marginal tax rate for the upper 1% of incomes (*TAX*). We are interested in whether the marginal tax rate is a useful instrument for reducing inequality. The resulting model is

$$SHARE_t = \alpha_1 + \alpha_2 TAX_t + \alpha_3 YEAR_t + \alpha_4 YEAR_t^2 + e_t$$

a. Estimate this equation using data for Canada. Obtain both conventional and HAC standard errors. Compare the 95% interval estimates for α_2 from each of the standard errors.

b. Use an LM test with a 5% significance level and three lagged residuals to test for autocorrelation in the errors of the equation estimated in part (a). What do you conclude about the use of HAC standard errors in part (a)?

c. Estimate a parameter ρ by applying OLS to the equation $\hat{e}_t = \rho\hat{e}_{t-1} + \hat{v}_t$ where $\hat{e}_t$ are the least squares residuals from part (a). What assumption is being made when you estimate this equation?

d. Transform each of the variables in the original equation using a transformation of the form $x_t^* = x_t - \hat{\rho}x_{t-1}$ and apply OLS to the transformed variables. Compute both conventional and HAC standard errors. Find the resulting 95% interval estimates for α_2. Compare them with each other and with those found in part (a).

e. Use an LM test with a 5% significance level and three lagged residuals to test for autocorrelation in the errors of the equation estimated in part (d). What do you conclude about the use of HAC standard errors in part (d)?

f. For each of the equations estimated in parts (a) and (d), discuss whether the exogeneity assumption required for consistent estimation of α_2 is likely to be satisfied.

9.28 In this exercise, we use a subset of the data compiled by Everaert and Pozzi[18] to forecast growth in per capita private consumption (*CONSN*) and growth in per capita real disposable income (*INC*) in France. Their data are annual from 1971 to 2007 and are stored in the data file *france_ep*.

a. To forecast consumption growth consider the autoregressive model

$$CONSN_t = \delta + \sum_{s=1}^{p} \theta_s CONSN_{t-s} + e_t$$

Estimate this model for $p = 1, 2, 3$, and 4. In each case, use 33 observations to ensure the same number of observations for each value of p. Based on significance of coefficients, autocorrelation in the residuals, and the Schwarz criterion, choose a suitable value for p.

b. For the choice of p in part (a), reestimate the model using all available observations and use it to find 95% interval forecasts for $CONSN_{2008}$, $CONSN_{2009}$ and $CONSN_{2010}$.

c. To forecast income growth, consider the ARDL model

$$INC_t = \delta + \sum_{s=1}^{p} \theta_s INC_{t-s} + \sum_{r=1}^{q} \delta_r HOURS_{t-r} + e_t$$

Estimate this model for $p = 1, 2$ and $q = 1, 2$. In each case, use 35 observations to ensure the same number of observations for all values of p and q. Use the Schwarz criterion to choose between the four models. In the model of your choice, are the coefficient estimates significantly different from zero at a 5% level? At a 10% level? Does the correlogram of residuals suggest that there is any serial correlation?

d. Use the model chosen in part (c) to find 95% interval forecasts for INC_{2008}, INC_{2009}, and INC_{2010}, given that $HOURS_{2008} = HOURS_{2009} = -0.0066$.

9.29 One way of modeling supply response for an agricultural crop is to specify a model in which area planted *AREA* depends on expected price, $PRICE^*$. A log-log (constant elasticity) version of this

[18]Everaert, G. and L. Ponzi (2014), "The Predictability of Aggregate Consumption Growth in OECD Countries: a Panel Data Analysis," *Journal of Applied Econometrics*, 29(3), 431–453.

model is $\ln(AREA_t) = \alpha + \gamma \ln(PRICE^*_{t+1}) + e_t$ where $PRICE^*_{t+1}$ is expected price in the next period when harvest takes place. When farmers expect price to be high, they plant more than when a low price is expected. Since they do not know the price at harvest time, we assume that they base their expectations on current and past prices, $\ln(PRICE^*_{t+1}) = \sum_{s=0}^{q} \gamma_s \ln(PRICE_{t-s})$, with more recent prices given a greater weight, $\gamma_0 > \gamma_1 > \cdots > \gamma_q$. 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 73 annual observations on area and price are given in the data file *bangla5*.

a. Let $\beta_s = \gamma\gamma_s$. Show that the model can be written as the finite distributed lag model

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

b. Estimate the model in part (a) assuming $q = 3$. Use HAC standard errors. What are the estimated delay and interim elasticities? Comment on the results. What are the first four autocorrelations of the residuals? Are they significantly different from zero at a 5% significance level?

c. You will have discovered that the lag weights obtained in part (a) do not satisfy a priori expectations. One way to try and overcome this problem is to insist that the weights lie on a straight line

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

If $\alpha_0 > 0$ and $\alpha_1 < 0$, these weights will decline implying farmers place a larger weight on more recent prices when forming their expectations. Substitute $\beta_s = \alpha_0 + \alpha_{1s}$ 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 $z_{t0} = \sum_{s=0}^{3} \ln(PRICE_{t-s})$ and $z_{t1} = \sum_{s=1}^{3} s \ln(PRICE_{t-s})$.

d. Create variables z_{t0} and z_{t1} and find least squares estimates of α_0 and α_1. Use HAC standard errors.

e. 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?

f. How do the delay and interim elasticities compare with those obtained earlier?

9.30 In this exercise, we consider a partial adjustment model as an alternative to the model used in Exercise 9.29 for modeling sugar cane area response in Bangladesh. The data are in the file *bangla5*. In the partial adjustment model long-run desired area, $AREA^*$ is a function of price,

$$AREA^*_t = \alpha + \beta_0 PRICE_t \qquad \text{(XR 9.30.1)}$$

In the short-run, fixed resource constraints prevent farmers from fully adjusting to the area desired at the prevailing price. Specifically,

$$AREA_t - AREA_{t-1} = \gamma(AREA^*_t - AREA_{t-1}) + e_t \qquad \text{(XR 9.30.2)}$$

where $AREA_t - AREA_{t-1}$ is the actual adjustment from the previous year, $AREA^*_t - AREA_{t-1}$ is the desired adjustment from the previous year, and $0 < \gamma < 1$.

a. Combine (XR 9.30.1) and (XR 9.30.2) to show that an estimable form of the model can be written as

$$AREA_t = \delta + \theta_1 AREA_{t-1} + \delta_0 PRICE_t + e_t$$

where $\delta = \alpha\gamma$, $\theta_1 = 1 - \gamma$, and $\delta_0 = \beta_0\gamma$.

b. Find least squares estimates of δ, θ_1, and δ_0. Are they significantly different from zero at a 5% significance level?

c. What are the first three autocorrelations of the residuals? Are they significantly different from zero at a 5% significance level?

d. Find estimates and standard errors for α, β_0, and γ. Are the estimates significantly different from zero at a 5% significance level?

e. Find an estimate of $AREA^*_{73}$ and compare it with $AREA_{73}$.

f. Forecast $AREA_{74}, AREA_{75}, \ldots, AREA_{80}$ assuming that price in the next 7 years does not change from the last sample value $\left(PRICE_{74} = PRICE_{75} = \cdots = PRICE_{80} = PRICE_{73}\right)$. Comment on these forecasts and compare the forecast $\widehat{AREA}_{80}$ with $AREA^*_{80}$ estimated from (XR 9.30.1).

9.31 Using data on the Maltese economy, Apap and Gravino[19] estimate a number of versions of Okun's Law. Their quarterly data run from 1999Q1 to 2012Q4 and can be found in the data file *apap*. The variables used in this exercise are $DU_t = U_t - U_{t-4}$ (the change in the unemployment rate relative to the same quarter in the previous year) and G_t (real output growth in quarter t relative to quarter $t-4$).

a. Estimate the Okun's Law equation $DU_t = \alpha + \beta_0 G_t + e_t$. Find both conventional and HAC standard errors and comment on the results.

b. Check the correlogram of the residuals $\hat{e}_t$ from the equation estimated in part (a). Is there evidence of autocorrelation?

c. Create the variable $q_t = G_t \times \hat{e}_t$, and examine its correlogram. Use this correlogram and equation (9.63) to suggest a reason why the conventional and HAC standard errors for the estimate of β_0 are similar in magnitude.

d. Estimate the finite distributed lag model

$$DU_t = \alpha + \beta_0 G_t + \beta_1 G_{t-1} + \beta_2 G_{t-2} + e_t$$

Use HAC standard errors. Is there evidence of a lagged effect of growth on unemployment? Using HAC standard errors in both cases, find a 95% interval estimate for the total multiplier and compare it with a 95% interval for the total multiplier from the model in part (a).

e. Estimate ARDL models $DU_t = \delta + \sum_{s=1}^{p} \theta_s DU_{t-s} + \sum_{r=0}^{q} \delta_r G_{t-r} + e_t$ for $p = 1, 2, 3$ and $q = 0, 1, 2$. Use HAC standard errors. Select and report the model with the largest number of lags whose coefficients are significantly different from zero at a 5% level.

f. For the model selected in part (e), find estimates for the total multiplier, the impact multiplier, and the first three delay multipliers of the infinite distributed lag representation.

g. For the model selected in part (e), find 95% interval estimates for the total multiplier and the two-period interim multiplier. How do they compare with the interval obtained in part (d)?

9.32 In their paper referred to in Exercise 9.31, Apap and Gravino examine the separate effects of output growth in the manufacturing and services sectors on changes in the unemployment rate. Their quarterly data run from 1999Q1 to 2012Q4 and can be found in the data file *apap*. The variables used in this exercise are $DU_t = U_t - U_{t-4}$ (the change in the unemployment rate relative to the same quarter in the previous year), MAN_t (real output growth in the manufacturing sector in quarter t relative to quarter $t-4$), SER_t (real output growth in the services sector in quarter t relative to quarter $t-4$), MAN_WT_t (the proportion of real output attributable to the manufacturing sector in quarter t), and SER_WT_t (the proportion of real output attributable to the services sector in quarter t). The relative effects of growth in each of the sectors on unemployment will depend not only on their growth rates but also on the relative size of each sector in the economy. To recognize this fact, construct the weighted growth variables $MAN2_t = MAN_t \times MAN\_WT_t$ and $SER2_t = SER_t \times SER\_WT_t$.

a. Use OLS with HAC standard errors to estimate the model

$$DU_t = \alpha + \gamma_0 SER2_t + \gamma_1 SER2_{t-1} + \beta_0 MAN2_t + \beta_1 MAN2_{t-1} + v_t$$

Comment on the relative importance of growth in each sector on changes in unemployment and on whether there is a lag in the effect from each sector.

b. Use an LM test with two lags and a 5% significance level to test for autocorrelation in the errors for the equation in part (a).

c. Assume that the errors in the equation in part (a) follow the AR(1) process $e_t = \rho e_{t-1} + v_t$. Show that, under this assumption, the model can be written as

$$\begin{aligned} DU_t = {} & \alpha(1-\rho) + \rho DU_{t-1} + \gamma_0 SER2_t + \left(\gamma_1 - \rho\gamma_0\right) SER2_{t-1} - \rho\gamma_1 SER2_{t-2} \\ & + \beta_0 MAN2_t + \left(\beta_1 - \rho\beta_0\right) MAN2_{t-1} - \rho\beta_1 MAN2_{t-2} + v_t \end{aligned}$$

d. Use nonlinear least squares with HAC standard errors to estimate the model in part (c). Have your conclusions made in part (a) changed?

[19] Apap, W. and D. Gravino (2017), "A Sectoral Approach to Okun's Law", *Applied Economics Letters* 25(5), 319–324. The authors are grateful to Wayne Apap for providing the data.

e. Use an LM test with two lags and a 5% significance level to test for autocorrelation in the errors for the equation in part (d). Is the AR(1) process adequate to model the autocorrelation in the errors of the original equation.

f. Suppose that, wanting to forecast DU_{2013Q1} using current and past information, you set up the model

$$DU_t = \delta + \theta_1 DU_{t-1} + \theta_2 DU_{t-2} + \gamma_1 SER2_{t-1} + \gamma_2 SER2_{t-2} + \delta_1 MAN2_{t-1} + \delta_2 MAN2_{t-2} + v_t$$

i. Have a sufficient number of lags of *DU* been included?

ii. Using a 5% significance level, test whether *SER2* Granger causes *DU*.

iii. Using a 5% significance level, test whether *MAN2* Granger causes *DU*.

9.33 The data file *xrate* contains monthly observations from 1986M1 to 2008M12 on the following variables[20]:

NER = the nominal exchange rate for the Australian dollar in terms of U.S. cents.
INF_AUS = the Australian inflation rate.
INF_US = the U.S. inflation rate.
DI6_AUS = the percentage change in the interest rate on an Australian government debt instrument of maturity 6 months.
DI6_US = the percentage change in the interest rate on a U.S. government debt instrument of maturity 6 months.

a. Plot *NER* against time and examine its correlogram. Does the series wander like a nonstationary series? Do the autocorrelations die out relatively quickly, suggesting a weakly dependent series?

b. Construct a variable which is the monthly change in the exchange rate, $DNER_t = NER_t - NER_{t-1}$. Plot *DNER* against time and examine its correlogram. Does the series wander like a nonstationary series? Do the autocorrelations die out relatively quickly, suggesting a weakly dependent series?

c. Theory suggests that the exchange rate will be higher when Australian inflation is low relative to that in the United States, and when the Australian interest rate is high relative to the U.S. interest rate. Construct the two variables $DINF_t = INF\_AUS_t - INF\_US_t$ and $DI6_t = DI6\_AUS_t - DI6\_US_t$, and estimate the model (using HAC standard errors)

$$DNER_t = \alpha + \beta_0 DINF_t + \beta_1 DINF_{t-1} + \gamma_0 DI6_t + \gamma_1 DI6_{t-1} + e_t$$

Comment on the results. Do the coefficients have the expected signs? Are they significantly different from zero using one-tail tests and a 5% significance level?

d. Reestimate the model in part (c), dropping variables whose coefficients had the wrong sign. Are the coefficients in the reestimated model significantly different from zero using one-tail tests and a 5% significance level? Check for serial correlation in the errors, using both the residual correlogram and an LM test with one lagged residual.

e. Reestimate the model in part (d) using feasible generalized least squares and assuming AR(1) errors. Estimate the model with both conventional and HAC standard errors. Are the coefficients in the reestimated model significantly different from zero using one-tail tests and a 5% significance level?

f. Suppose that the following model is proposed for 1-month ahead forecasting of the exchange rate

$$DNER_t = \delta + \theta_1 DNER_{t-1} + \delta_1 DINF_{t-1} + \phi_1 DI6_{t-1} + e_t$$

Estimate this model using observations from 1986M1 to 2007M12. Does it appear to be a good model for forecasting?

g. Use the model in part (f) to obtain 1-month ahead forecasts of *NER* for each of the months in 2008. (Use the actual values of $DNER_{t-1}$ to obtain each forecast.) Comment on the accuracy of the forecasts and compute the average absolute forecast error $\sum_{t=2008M1}^{2008M12} \left| \widehat{NER}_t - NER_t \right| \Big/ 12$.

9.34 In the new Keynesian Phillips curve (NKPC), inflation at time t (INF_t) depends on inflationary expectations formed at time t for time $t+1$ $(INFEX_t)$, and the output gap, defined as output less potential output. Expectations of higher inflation lead to greater inflation. The closer output is to potential

[20]These data are constructed from the data archive for Berge, T. (2014), "Forecasting Disconnected Exchange Rates," *Journal of Applied Econometrics* 29(5), 713–735.

output, the higher the inflation rate. Amberger et al.[21] compare results from estimating NKPCs with two output gaps, one that has been augmented with changes in financial variables $\left(FNGAP_t\right)$, and one that has not $\left(GAP_t\right)$. Quarterly data for Italy for the period 1990Q1 to 2014Q4 can be found in the data file *italy*.

a. Using OLS, estimate the two equations

$$INF_t = \alpha_G + \beta_G INFEX_t + \gamma_G GAP_t + e_{Gt}$$
$$INF_t = \alpha_F + \beta_F INFEX_t + \gamma_F FNGAP_t + e_{Ft}$$

Find 95% interval estimates for γ_G and γ_F using both conventional and HAC standard errors. Comment on (i) the relative widths of the intervals with and without HAC standard errors and (ii) whether one output gap measure is preferred over another in terms of its impact on inflation.

b. What are the values of the first four residual autocorrelations from each of the two regressions in part (a)? Which ones are significantly different from zero at a 5% significance level?

c. Consider the generic equation $y_t = \alpha + \beta x_t + \gamma z_t + e_t$ with AR(2) errors $e_t = \psi_1 e_{t-1} + \psi_2 e_{t-2} + v_t$ where the v_t are not autocorrelated. Show that this model can be written as

$$y_t^* = \alpha^* + \beta x_t^* + \gamma z_t^* + v_t \qquad t = 3, 4, \ldots, T$$

where $y_t^* = y_t - \psi_1 y_{t-1} - \psi_2 y_{t-2}$, $\alpha^* = \alpha\left(1 - \psi_1 - \psi_2\right)$, $x_t^* = x_t - \psi_1 x_{t-1} - \psi_2 x_{t-2}$, and $z_t^* = z_t - \psi_1 z_{t-1} - \psi_2 z_{t-2}$.

d. Using the least squares residuals $\hat{e}_{Gt}$ from the first equation in part (a), estimate ψ_1 and ψ_2 from the regression equation $\hat{e}_{Gt} = \psi_1 \hat{e}_{G,t-1} + \psi_2 \hat{e}_{G,t-2} + \hat{v}_t$. Along the lines of the transformations in part (c), use the estimates of ψ_1 and ψ_2 to find transformed variables INF_t^*, $INFEX_t^*$, and GAP_t^* and then estimate α_G^*, β_G, and γ_G from the transformed equation $INF_t^* = \alpha_G^* + \beta_G INFEX_t^* + \gamma_G GAP_t^* + v_t$. Estimate the equation with both conventional and HAC standard errors.

e. Using the results from part (d), find 95% interval estimates for γ_G using both conventional and HAC standard errors. Comment on (i) the relative widths of the intervals with and without HAC standard errors and (ii) how the estimates and intervals compare with the corresponding ones obtained in part (a).

9.35 Do lags of the variables in the new Keynesian Phillips curve provide a good basis for forecasting quarterly inflation? In this exercise, we investigate this question using the French data from Amberger et al. See Exercise 9.34 for details. The data are stored in the data file *france*.

a. Consider ARDL models of the form

$$INF_t = \delta + \sum_{s=1}^{p} \theta_s INF_{t-s} + \sum_{r=1}^{q} \delta_r INFEX_{t-r} + \sum_{j=1}^{m} \gamma_j GAP_{t-j} + e_t$$

Using observations from 1991Q1 to 2013Q4, estimate this equation for $p = 2$, $q = 1, 2, 3, 4$ and $m = 1, 2, 3, 4$. From these 16 equations, select and report the one with the smallest value of the Schwarz criterion. Note that 92 observations should be used to estimate each equation.

b. In the equation selected in part (a), are all the estimated coefficients significantly different from zero at a 5% significance level? Does the correlogram suggest that there is no autocorrelation in the errors?

c. Use the selected model from part (a) to find 95% forecast intervals for inflation in 2014Q1, 2014Q2, 2014Q3, and 2014Q4. When computing the forecasts, use actual values of *INFEX* and *GAP* where needed but assume that the actual values of *INF* in the four forecast quarters are unknown. After you have found the forecast intervals, check whether the actual values lie within those intervals. [*Hint*: If your software does not compute standard errors of forecast errors, equation (9.41) can be used to find them for the first three quarters. For the fourth quarter, the variance of the forecast error is given by

$$\sigma_{f4}^2 = \left[\left(\theta_1^3 + 2\theta_1\theta_2\right)^2 + \left(\theta_1^2 + \theta_2\right)^2 + \theta_1^2 + 1\right]\sigma^2$$

You might like to prove this result.]

d. What assumptions are necessary for the standard errors of the forecast errors to be valid?

---

[21] Amberger, J., R Fendel and H. Stremmel (2017), "Improved output gaps with financial cycle information? An application to G7 countries' new Keynesian Phillips curves," *Applied Economics Letters*, 24(4), 219–228. Many thanks to Johanna Amberger for supplying the data used in this study.

9.36 Consider the following model where a dependent variable y depends on infinite distributed lags of the two variables x and z.

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

Suppose that both sets of lag weights decline geometrically, but with different parameters λ_1 and λ_2. That is, $\beta_s = \lambda_1^s \beta_0$ and $\gamma_r = \lambda_2^r \gamma_0$.

a. Show that the model can be written as

$$y_t = \alpha + \beta_0 \sum_{s=0}^{\infty} \lambda_1^s L^s x_t + \gamma_0 \sum_{r=0}^{\infty} \lambda_2^r L^r z_t + e_t$$

b. Use the result in Exercise 9.15 to show that the equation in (a) can be written as

$$\begin{aligned} y_t &= \alpha + \beta_0 \left(1 - \lambda_1 L\right)^{-1} x_t + \gamma_0 \left(1 - \lambda_2 L\right)^{-1} z_t + e_t \\ &= \alpha^* + \left(\lambda_1 + \lambda_2\right) y_{t-1} - \lambda_1 \lambda_2 y_{t-2} + \beta_0 x_t - \beta_0 \lambda_2 x_{t-1} + \gamma_0 z_t - \gamma_0 \lambda_1 z_{t-1} + v_t \end{aligned}$$

where $\alpha^* = \left(1 - \lambda_1\right)\left(1 - \lambda_2\right)\alpha$ and $v_t = e_t - \left(\lambda_1 + \lambda_2\right) e_{t-1} + \lambda_1 \lambda_2 e_{t-2}$.

c. Using data in the file *canada5*, with $y_t = INF_t$, $x_t = INFEX_t$, and $z_t = GAP_t$, estimate the last equation in part (b) using nonlinear least squares. Report the estimates, their standard errors, and one-tail p-values for a zero null hypothesis on each parameter (except the constant). Are the estimates significantly different from zero at a 5% level?

d. Find estimates of the first three lag weights for both *INFEX* and *GAP*.

e. Find estimates of the total multipliers for both *INFEX* and *GAP*.

f. Using a 5% significance level, test $H_0 : \lambda_1 = \lambda_2$ versus $H_1 : \lambda_1 \neq \lambda_2$. What are the implications for the model if H_0 is true?

g. The equation estimated in part (c) can be viewed as a restricted version of the more general ARDL(2, 1, 1) model

$$y_t = \alpha^* + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \delta_0 x_t + \delta_1 x_{t-1} + \phi_0 z_t + \phi_1 z_{t-1} + v_t$$

where $\dfrac{\delta_1}{\delta_0} \times \dfrac{\phi_1}{\phi_0} = -\theta_2$ and $\dfrac{\delta_1}{\delta_0} + \dfrac{\phi_1}{\phi_0} = -\theta_1$. Estimate this unrestricted model and jointly test the validity of the restrictions at a 5% level. What are the implications for the infinite distributed lags if the restrictions are not true?

h. Test the hypothesis that e_t follows an AR(2) process $e_t = \left(\lambda_1 + \lambda_2\right) e_{t-1} - \lambda_1 \lambda_2 e_{t-2} + u_t$. What are the implications of rejecting this hypothesis?

Appendix 9A The Durbin–Watson Test

In Section 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. In addition, the test is derived conditional on $\mathbf{X}$; it treats the explanatory variables as nonrandom.

It is assumed that the v_t are independent random errors with distribution $N\left(0, \sigma_v^2\right)$ 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

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

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

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

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

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

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 $d \le d_c$? 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 $P\left(d \le d_c\right) = 0.05$. Then, as illustrated in Figure 9.A1, we reject H_0 if $d \le d_c$ 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 the calculated value of d. Thus, if the p-value is less than or equal to 0.05, it follows that $d \le d_c$, and H_0 is rejected. If the p-value is greater than 0.05, then $d > d_c$, and H_0 is not rejected.

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

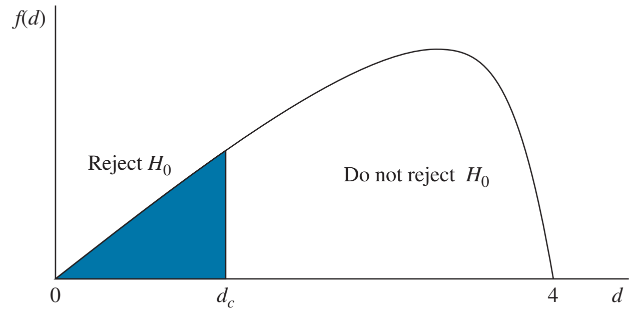

FIGURE 9.A1 **Testing for positive autocorrelation.**

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.[1]

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

$$d_L < d < d_U$$

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 9.A2. Let d_{Lc} be the 5% critical value from the probability distribution for d_L. That is, d_{Lc} is such that $P(d_L \leq d_{Lc}) = 0.05$. Similarly, let d_{Uc} be such that $P(d_U \leq d_{Uc}) = 0.05$. 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 9.A2, 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 $d \leq d_{Lc}$, then it must follow that $d \leq d_c$, and H_0 is rejected. In addition, 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*:

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

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.

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

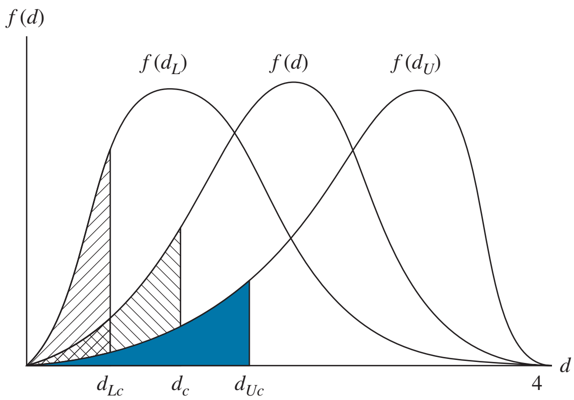

FIGURE 9.A2 **Upper and lower critical value bounds for the Durbin–Watson test.**

EXAMPLE 9.20 | Durbin–Watson Bounds Test for Phillips Curve

The 5% critical bounds for the Phillips curve in Examples 9.14 and 9.15, for $T = 117$ and $K = 2$ are[2]

$$d_{Lc} = 1.681 \qquad d_{Uc} = 1.716$$

The Durbin–Watson test value is 0.965. Since $0.965 < d_{Lc} = 1.681$, we conclude that $d < d_c$, and hence we reject $H_0: \rho = 0$; there is evidence to suggest that the errors are positively serially correlated.

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.[3] 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

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

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

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

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

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

---

[2]These values can be found from the Durbin Watson tables on the web site principlesofeconometrics.com/poe5/poe5.htm.

[3]To simplify the exposition, we derive these properties in terms of the marginal distributions of e_t and v_t. When estimating the AR(1) error model in the body of the chapter, we make the stronger assumptions $E\left(v_t|\mathbf{X}\right) = 0$ and $\text{var}\left(v_t|\mathbf{X}\right) = \sigma_v^2$.

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

$$\begin{aligned} \text{var}(e_t) &= \text{var}(v_t) + \rho^2 \text{var}(v_{t-1}) + \rho^4 \text{var}(v_{t-2}) + \rho^6 \text{var}(v_{t-3}) + \cdots \\ &= \sigma_v^2 + \rho^2\sigma_v^2 + \rho^4\sigma_v^2 + \rho^6\sigma_v^2 + \cdots \\ &= \sigma_v^2 (1 + \rho^2 + \rho^4 + \rho^6 + \cdots) \\ &= \frac{\sigma_v^2}{1-\rho^2} \end{aligned} \tag{9B.5}$$

In the abovementioned 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

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

When the second line in the abovementioned 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 $E(v_{t-k}^2) = \text{var}(v_{t-k}) = \sigma_v^2$ for all lags k. In a similar way, we can show that the covariance between errors that are k periods apart is

$$\text{cov}(e_t, e_{t-k}) = \frac{\rho^k \sigma_v^2}{1-\rho^2} \qquad k > 0 \tag{9B.6}$$

From (9B.5) and (9B.6), the autocorrelations for errors that are k periods apart are given by

$$\rho_k = \text{corr}(e_t, e_{t-k}) = \frac{\text{cov}(e_t, e_{t-k})}{\text{var}(e_t)} = \frac{\rho^k \sigma_v^2/(1-\rho^2)}{\sigma_v^2/(1-\rho^2)} = \rho^k$$