CHAPTER 1

DIFFERENCE EQUATIONS

INTRODUCTION

The theory of difference equations underlies all of the time-series methods employed in later chapters of this text. It is fair to say that time-series econometrics is concerned with the estimation of difference equations containing stochastic components. The traditional use of time-series analysis was to forecast the time path of a variable. Uncovering the dynamic path of a series improves forecasts since the predictable components of the series can be extrapolated into the future. The growing interest in economic dynamics has given a new emphasis to time-series econometrics. Stochastic difference equations arise quite naturally from dynamic economic models. Appropriately estimated equations can be used for the interpretation of economic data and for hypothesis testing.

This introductory chapter has three aims:

1. Explain how stochastic difference equations can be used for forecasting and illustrate how such equations can arise from familiar economic models. The chapter is not meant to be a treatise on the theory of difference equations. Only those techniques that are essential to the appropriate estimation of *linear* time-series models are presented. This chapter focuses on single equation models; multivariate models are considered in Chapters 5 and 6.
2. Explain what it means to *solve* a difference equation. The solution will determine whether a variable has a stable or an explosive time path. Knowledge of the stability conditions is essential to understanding the recent innovations in time-series econometrics. The contemporary time-series literature pays special attention to the issue of stationary versus nonstationary variables. The stability conditions underlie the conditions for stationarity.
3. Demonstrate how to find the solution to a stochastic difference equation. There are several different techniques that can be used; each has its own relative merits. A number of examples are presented to help you understand the different methods. Try to work through each example carefully. For extra practice, you should answer the exercises at the end of the chapter.

1. TIME-SERIES MODELS

The task facing the modern time-series econometrician is to develop reasonably simple models capable of forecasting, interpreting, and testing hypotheses concerning economic data. The challenge has grown over time; the original use of time-series analysis was primarily as an aid to forecasting. As such, a methodology was developed to decompose a series into a trend, a seasonal, a cyclical, and an irregular component. Uncovering the dynamic path of a series improves forecast accuracy because each of the predictable components can be extrapolated into the future.

Suppose you observe the fifty data points shown in Figure 1.1 and are interested in forecasting the subsequent values. Using the time-series methods discussed in the next several chapters, it is possible to decompose this series into the trend, seasonal, and irregular components shown in the lower panel of the figure. As you can see, the trend changes the mean of the series, and the seasonal component imparts a regular cyclical pattern with peaks occurring every twelve units of time. In practice, the trend and seasonal components will not be the simplistic deterministic functions shown in this figure. With economic data, it is typical to find that a series contains stochastic elements in the trend, seasonal, and irregular components. For the time being, it is wise to sidestep these complications so that the projection of the trend and seasonal components into periods 51 and beyond is straightforward.

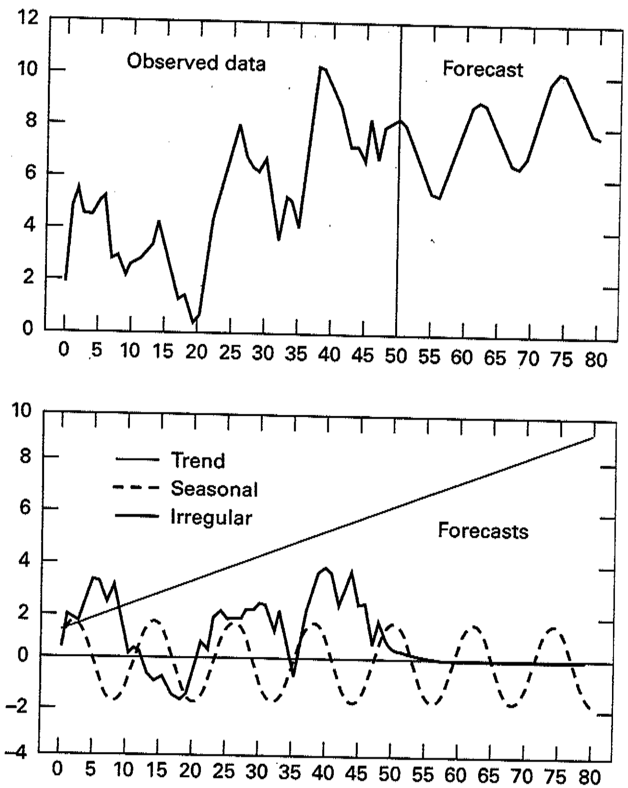


FIGURE 1.1 Hypothetical Time Series

Notice that the irregular component, while lacking a well-defined pattern, is somewhat predictable. If you examine the figure closely, you will see that the positive and negative values occur in runs; the occurrence of a large value in any period tends to be followed by another large value. Short-run forecasts will make use of this positive correlation in the irregular component. Over the entire span, however, the irregular component exhibits a tendency to revert to zero. As shown in the lower part, the projection of the irregular component past period 50 rapidly decays toward zero. The overall forecast, shown in the top part of the figure, is the sum of each forecasted component.

The general methodology used to make such forecasts entails finding the *equation of motion* driving a stochastic process and using that equation to predict subsequent outcomes. Let y_t denote the value of a data point at period t; if we use this notation, the example in Figure 1.1 assumes we observed y_1 through y_{50}. For $t = 1$ to 50, the equations of motion used to construct components of the y_t series are

$$\text{Trend: } T_t = 1 + 0.1t$$
$$\text{Seasonal: } S_t = 1.6 \sin(t\pi/6)$$
$$\text{Irregular: } I_t = 0.7\, I_{t-1} + \varepsilon_t$$

where: T_t = value of the trend component in period t
S_t = value of the seasonal component in t
I_t = the value of the irregular component in t
ε_t = a pure random disturbance in t

Thus, the irregular disturbance in t is 70 percent of the previous period's irregular disturbance plus a random disturbance term.

Each of these three equations is a type of **difference equation.** In its most general form, a difference equation expresses the value of a variable as a function of its own lagged values, time, and other variables. The trend and seasonal terms are both functions of time and the irregular term is a function of its own lagged value and of the stochastic variable ε_t. The reason for introducing this set of equations is to make the point that *time-series econometrics is concerned with the estimation of difference equations containing stochastic components.* The time-series econometrician may estimate the properties of a single series or a vector containing many interdependent series. Both univariate and multivariate forecasting methods are presented in the text. Chapter 2 shows how to estimate the irregular part of a series. Chapter 3 considers estimating the variance when the data exhibit periods of volatility and tranquility. Estimation of the trend is considered in Chapter 4, which focuses on the issue of whether the trend is deterministic or stochastic. Chapter 5 discusses the properties of a vector of stochastic difference equations, and Chapter 6 is concerned with the estimation of trends in a multivariate model.

Although forecasting has always been the mainstay of time-series analysis, the growing importance of economic dynamics has generated new uses for time-series analysis. Many economic theories have natural representations as stochastic difference equations. Moreover, many of these models have testable implications concerning the time path of a key economic variable. Consider the following three examples:

1. **The Random Walk Hypothesis:** In its simplest form, the random walk model suggests that day-to-day changes in the price of a stock should have

a mean value of zero. After all, if it is known that a capital gain can be made by buying a share on day t and selling it for an expected profit the very next day, efficient speculation will drive up the current price. Similarly, no one will want to hold a stock if it is expected to depreciate. Formally, the model asserts that the price of a stock should evolve according to the stochastic difference equation

$$y_{t+1} = y_t + \varepsilon_{t+1}$$

or

$$\Delta y_{t+1} = \varepsilon_{t+1}$$

where: y_t = the price of a share of stock on day t
ε_{t+1} = a random disturbance term that has an expected value of zero

Now consider the more general stochastic difference equation:

$$\Delta y_{t+1} = \alpha_0 + \alpha_1 y_t + \varepsilon_{t+1}$$

The random walk hypothesis requires the testable restriction: $\alpha_0 = \alpha_1 = 0$. Rejecting this restriction is equivalent to rejecting the theory. Given the information available in period t, the theory also requires that the mean of ε_{t+1} be equal to zero; evidence that ε_{t+1} is predictable invalidates the random walk hypothesis. Again, the appropriate estimation of a single equation model is considered in Chapters 2 through 4.

2. **Reduced-Form and Structural Equations:** Often it is useful to collapse a system of difference equations into separate single-equation models. To illustrate the key issues involved, consider a stochastic version of Samuelson's (1939) classic model:

$$y_t = c_t + i_t \qquad (1.1)$$

$$c_t = \alpha y_{t-1} + \varepsilon_{ct} \qquad 0 < \alpha < 1 \qquad (1.2)$$

$$i_t = \beta(c_t - c_{t-1}) + \varepsilon_{it} \qquad \beta > 0 \qquad (1.3)$$

where y_t, c_t, and i_t denote real GDP, consumption, and investment in time period t, respectively. In this Keynesian model, y_t, c_t, and i_t are endogenous variables. The previous period's GDP and consumption, y_{t-1} and c_{t-1}, are called predetermined or lagged endogenous variables. The terms ε_{ct} and ε_{it} are zero mean random disturbances for consumption and investment, and the coefficients α and β are parameters to be estimated.

The first equation equates aggregate output (GDP) with the sum of consumption and investment spending. The second equation asserts that consumption spending is proportional to the previous period's GDP plus a random disturbance term. The third equation illustrates the accelerator principle. Investment spending is proportional to the change in consumption; the idea is that growth in consumption necessitates new investment spending. The error terms ε_{ct} and ε_{it} represent the portions of consumption and investment not explained by the behavioral equations of the model.

Equation (1.3) is a **structural equation** since it expresses the endogenous variable i_t as being dependent on the current realization of another endogenous variable, c_t. A **reduced-form equation** is one expressing the value of a variable in terms of its own lags, lags of other endogenous variables, current and past values of exogenous variables, and disturbance terms. As formulated, the consumption function is already in reduced form; current consumption depends only on lagged income and the current value of the stochastic disturbance term ε_{ct}. Investment is not in reduced form because it depends on current period consumption.

To derive a reduced-form equation for investment, substitute (1.2) into the investment equation to obtain

$$\begin{aligned} i_t &= \beta[\alpha y_{t-1} + \varepsilon_{ct} - c_{t-1}] + \varepsilon_{it} \\ &= \alpha\beta y_{t-1} - \beta c_{t-1} + \beta\varepsilon_{ct} + \varepsilon_{it} \end{aligned}$$

Notice that the reduced-form equation for investment is not unique. You can lag (1.2) one period to obtain: $c_{t-1} = \alpha y_{t-2} + \varepsilon_{ct-1}$. Using this expression, the reduced-form investment equation can also be written as

$$\begin{aligned} i_t &= \alpha\beta y_{t-1} - \beta(\alpha y_{t-2} + \varepsilon_{ct-1}) + \beta\varepsilon_{ct} + \varepsilon_{it} \\ &= \alpha\beta(y_{t-1} - y_{t-2}) + \beta(\varepsilon_{ct} - \varepsilon_{ct-1}) + \varepsilon_{it} \end{aligned} \qquad (1.4)$$

Similarly, a reduced-form equation for GDP can be obtained by substituting (1.2) and (1.4) into (1.1):

$$\begin{aligned} y_t &= \alpha y_{t-1} + \varepsilon_{ct} + \alpha\beta(y_{t-1} - y_{t-2}) + \beta(\varepsilon_{ct} - \varepsilon_{ct-1}) + \varepsilon_{it} \\ &= \alpha(1+\beta)y_{t-1} - \alpha\beta y_{t-2} + (1+\beta)\varepsilon_{ct} + \varepsilon_{it} - \beta\varepsilon_{ct-1} \end{aligned} \qquad (1.5)$$

Equation (1.5) is a **univariate** reduced-form equation; y_t is expressed solely as a function of its own lags and disturbance terms. A univariate model is particularly useful for forecasting since it enables you to predict a series based solely on its own current and past realizations. It is possible to estimate (1.5) using the univariate time-series techniques explained in Chapters 2 through 4. Once you have obtained estimates of α and β, it is straightforward to use the observed values of y_1 through y_t to predict all future values in the series (i.e., $y_{t+1}, y_{t+2}, \ldots$).

Chapter 5 considers the estimation of multivariate models when all variables are treated as jointly endogenous. The chapter also discusses the restrictions needed to recover (i.e., identify) the structural model from the estimated reduced-form model.

3. **Error-Correction: Forward and Spot Prices:** Certain commodities and financial instruments can be bought and sold on the spot market (for immediate delivery) or for delivery at some specified future date. For example, suppose that the price of a particular foreign currency on the spot market is s_t dollars and that the price of the currency for delivery one period into the future is f_t dollars. Now, consider a speculator who purchased forward currency at the price f_t dollars per unit. At the beginning of period $t + 1$, the speculator receives the currency and pays f_t dollars per unit

received. Since spot foreign exchange can be sold at s_{t+1}, the speculator can earn a profit (or loss) of $s_{t+1} - f_t$ per unit transacted.

The Unbiased Forward Rate (UFR) hypothesis asserts that expected profits from such speculative behavior should be zero. Formally, the hypothesis posits the following relationship between forward and spot exchange rates:

$$s_{t+1} = f_t + \varepsilon_{t+1} \qquad (1.6)$$

where ε_{t+1} has a mean value of zero from the perspective of time period t.

In (1.6), the forward rate in t is an unbiased estimate of the spot rate in t+1. Thus, suppose you collected data on the two rates and estimated the regression

$$s_{t+1} = \alpha_0 + \alpha_1 f_t + \varepsilon_{t+1}$$

If you were able to conclude that $\alpha_0 = 0$, $\alpha_1 = 1$, and that the regression residuals ε_{t+1} have a mean value of zero from the perspective of time period t, the UFR hypothesis could be maintained.

The spot and forward markets are said to be in *long-run equilibrium* when $\varepsilon_{t+1} = 0$. Whenever s_{t+1} turns out to differ from f_t, some sort of adjustment must occur to restore the equilibrium in the subsequent period. Consider the adjustment process

$$s_{t+2} = s_{t+1} - \alpha[s_{t+1} - f_t] + \varepsilon_{st+2} \qquad \alpha > 0 \qquad (1.7)$$

$$f_{t+1} = f_t + \beta[s_{t+1} - f_t] + \varepsilon_{ft+1} \qquad \beta > 0 \qquad (1.8)$$

where ε_{st+2} and ε_{ft+1} both have a mean value of zero.

Equations (1.7) and (1.8) illustrate the type of simultaneous adjustment mechanism considered in Chapter 6. This dynamic model is called an **error-correction** model because the movement of the variables in any period is related to the previous period's gap from long-run equilibrium. If the spot rate s_{t+1} turns out to equal the forward rate f_t, (1.7) and (1.8) state that the spot rate and forward rates are expected to remain unchanged. If there is a positive gap between the spot and forward rates so that $s_{t+1} - f_t > 0$, (1.7) and (1.8) lead to the prediction that the spot rate will fall and the forward rate will rise.

2. DIFFERENCE EQUATIONS AND THEIR SOLUTIONS

Although many of the ideas in the previous section were probably familiar to you, it is necessary to formalize some of the concepts used. In this section, we will examine the type of difference equation used in econometric analysis and make explicit what it means to "solve" such equations. To begin our examination of difference equations, consider the function $y = f(t)$. If we evaluate the function when the independent variable t takes on the specific value t^*, we get a specific value for the dependent variable called y_{t^*}. Formally, $y_{t^*} = f(t^*)$. Using this same notation, y_{t^*+h} represents the value of y when t takes on the specific value $t^* + h$. The first difference of y is defined as

the value of the function when evaluated at $t = t^* + h$ minus the value of the function evaluated at t^*:

$$\Delta y_{t^*+h} \equiv f(t^* + h) - f(t^*)$$
$$\equiv y_{t^*+h} - y_{t^*} \qquad (1.9)$$

Differential calculus allows the change in the independent variable (i.e., the term h) to approach zero. Since most economic data is collected over discrete periods, however, it is more useful to allow the length of the time period to be greater than zero. Using difference equations, we normalize units so that h represents a unit change in t (i.e., $h = 1$) and consider the sequence of equally spaced values of the independent variable. Without any loss of generality, we can always drop the asterisk on t^*. We can then form the **first differences:**

$$\Delta y_t = f(t) - f(t-1) \equiv y_t - y_{t-1}$$
$$\Delta y_{t+1} = f(t+1) - f(t) \equiv y_{t+1} - y_t$$
$$\Delta y_{t+2} = f(t+2) - f(t+1) \equiv y_{t+2} - y_{t+1}$$

Often it will be convenient to express the entire sequence of values $\{\ldots y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, \ldots\}$ as $\{y_t\}$. We can then refer to any particular value in the sequence as y_t. Unless specified, the index t runs from $-\infty$ to $+\infty$. In time-series econometric models, we use t to represent "time" and h to represent the length of a time period. Thus, y_t and y_{t+1} might represent the realizations of the $\{y_t\}$ sequence in the first and second quarters of 2004, respectively.

In the same way we can form the **second difference** as the change in the first difference. Consider

$$\Delta^2 y_t \equiv \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$
$$\Delta^2 y_{t+1} \equiv \Delta(\Delta y_{t+1}) = \Delta(y_{t+1} - y_t) = (y_{t+1} - y_t) - (y_t - y_{t-1}) = y_{t+1} - 2y_t + y_{t-1}$$

The nth difference (Δ^n) is defined analogously. At this point, we risk taking the theory of difference equations too far. As you will see, the need to use second differences rarely arises in time-series analysis. It is safe to say that third and higher-order differences are never used in applied work.

Since this text considers linear time-series methods, it is possible to examine only the special case of an nth-order linear difference equation with constant coefficients. The form for this special type of difference equation is given by

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t \qquad (1.10)$$

The order of the difference equation is given by the value of n. The equation is linear because all values of the dependent variable are raised to the first power. Economic theory may dictate instances in which the various a_i are functions of variables within the economy. However, as long as they do not depend on any of the values of y_t or x_t, we can regard them as parameters. The term x_t is called the **forcing process.** The form of the forcing process can be very general; x_t can be any function of time, current and lagged values of other variables, and/or stochastic disturbances. From an appropriate choice of the forcing process, we can obtain a wide variety of important macroeconomic models. Re-examine equation (1.5), the reduced-form equation for GDP. This equation is a second-order difference equation since y_t depends on y_{t-2}. The forcing process is the expression $(1+\beta)\varepsilon_{ct} + \varepsilon_{it} - \beta\varepsilon_{ct-1}$. You will note that (1.5) has no intercept term corresponding to the expression a_0 in (1.10).

An important special case for the $\{x_t\}$ sequence is

$$x_t = \sum_{i=0}^{\infty} \beta_i \, \varepsilon_{t-i}$$

where the β_i are constants (some of which can equal zero) and the individual elements of the sequence $\{\varepsilon_t\}$ are not functions of the y_t. At this point it is useful to allow the $\{\varepsilon_t\}$ sequence to be nothing more than a sequence of unspecified exogenous variables. For example, let $\{\varepsilon_t\}$ be a random error term and set $\beta_0 = 1$ and $\beta_1 = \beta_2 = \ldots = 0$; in this case, (1.10) becomes the autoregression equation:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_n y_{t-n} + \varepsilon_t$$

Let $n = 1$, $a_0 = 0$ and $a_1 = 1$ to obtain the random walk model. Notice that equation (1.10) can be written in terms of the **difference operator** (Δ). Subtracting y_{t-1} from (1.10), we obtain

$$y_t - y_{t-1} = a_0 + (a_1 - 1) y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + x_t$$

or defining $\gamma = (a_1 - 1)$, we get

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + x_t \qquad (1.11)$$

Clearly, equation (1.11) is simply a modified version of (1.10).

A **solution** to a difference equation expresses the value of y_t as a function of the elements of the $\{x_t\}$ sequence and t (and possibly some given values of the $\{y_t\}$ sequence called **initial conditions**). Examining (1.11) makes it clear that there is a strong analogy to integral calculus, where the problem is to find a primitive function from a given derivative. We seek to find the primitive function $f(t)$, given an equation expressed in the form of (1.10) or (1.11). Notice that a solution is a function rather than a number. The key property of a solution is that it satisfies the difference equation for all permissible values of t and $\{x_t\}$. Thus, the substitution of a solution into the difference equation must result in an identity. For example, consider the simple difference equation $\Delta y_t = 2$ (or $y_t = y_{t-1} + 2$). You can easily verify that a solution to this difference equation is $y_t = 2t + c$, where c is any arbitrary constant. By definition, if $2t + c$ is a solution, it must hold for all permissible values of t. Thus for period $t-1$, $y_{t-1} = 2(t-1) + c$. Now substitute the solution into the difference equation to form

$$2t + c \equiv 2(t-1) + c + 2 \qquad (1.12)$$

It is straightforward to carry out the algebra and verify that (1.12) is an identity. This simple example also illustrates that the solution to a difference equation need not be unique; there is a solution for any arbitrary value of c.

Another useful example is provided by the irregular term shown in Figure 1.1; recall that the equation for this expression is: $I_t = 0.7I_{t-1} + \varepsilon_t$. You can verify that the solution to this first-order equation is

$$I_t = \sum_{i=0}^{\infty} (0.7)^i \, \varepsilon_{t-i} \qquad (1.13)$$

Since (1.13) holds for all time periods, the value of the irregular component in $t - 1$ is given by

$$I_{t-1} = \sum_{i=0}^{\infty} (0.7)^i \, \varepsilon_{t-1-i} \qquad (1.14)$$

Now substitute (1.13) and (1.14) into $I_t = 0.7I_{t-1} + \varepsilon_t$ to obtain

$$\begin{aligned} &\varepsilon_t + 0.7\varepsilon_{t-1} + (0.7)^2\varepsilon_{t-2} + (0.7)^3\varepsilon_{t-3} + \ldots \\ &= 0.7[\varepsilon_{t-1} + 0.7\varepsilon_{t-2} + (0.7)^2\varepsilon_{t-3} + (0.7)^3\varepsilon_{t-4} + \ldots] + \varepsilon_t \end{aligned} \qquad (1.15)$$

The two sides of (1.15) are identical; this proves that (1.13) is a solution to the first-order stochastic difference equation $I_t = 0.7I_{t-1} + \varepsilon_t$. Be aware of the distinction between reduced-form equations and solutions. Since $I_t = 0.7I_{t-1} + \varepsilon_t$ holds for all values of t, it follows that $I_{t-1} = 0.7I_{t-2} + \varepsilon_{t-1}$. Combining these two equations yields

$$I_t = 0.7[0.7I_{t-2} + \varepsilon_{t-1}] + \varepsilon_t = 0.49I_{t-2} + 0.7\varepsilon_{t-1} + \varepsilon_t \qquad (1.16)$$

Equation (1.16) is a reduced-form equation since it expresses I_t in terms of its own lags and disturbance terms. However, (1.16) does not qualify as a solution because it contains the "unknown" value of I_{t-2}. To qualify as a solution, (1.16) must express I_t in terms of the elements x_t, t, and any given initial conditions.

3. SOLUTION BY ITERATION

The solution given by (1.13) was simply postulated. The remaining portions of this chapter develop the methods you can use to obtain such solutions. Each method has its own merits; knowing the most appropriate to use in a particular circumstance is a skill that comes only with practice. This section develops the method of iteration. Although iteration is the most cumbersome and time-intensive method, most people find it to be very intuitive.

If the value of y in some specific period is known, a direct method of solution is to iterate forward from that period to obtain the subsequent time path of the entire y sequence. Refer to this known value of y as the initial condition or the value of y in time period 0 (denoted by y_0). It is easiest to illustrate the iterative technique using the first-order difference equation:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \qquad (1.17)$$

Given the value of y_0, it follows that y_1 will be given by

$$y_1 = a_0 + a_1 y_0 + \varepsilon_1$$

In the same way, y_2 must be

$$\begin{aligned} y_2 &= a_0 + a_1 y_1 + \varepsilon_2 \\ &= a_0 + a_1[a_0 + a_1 y_0 + \varepsilon_1] + \varepsilon_2 \\ &= a_0 + a_0 a_1 + (a_1)^2 y_0 + a_1\varepsilon_1 + \varepsilon_2 \end{aligned}$$

Continuing the process in order to find y_3, we obtain

$$y_3 = a_0 + a_1 y_2 + \varepsilon_3 = a_0[1 + a_1 + (a_1)^2] + (a_1)^3 y_0 + a_1^2\varepsilon_1 + a_1\varepsilon_2 + \varepsilon_3$$

You can easily verify that for all $t > 0$, repeated iteration yields

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \qquad (1.18)$$

Equation (1.18) is a solution to (1.17) since it expresses y_t as a function of t, the forcing process $x_t = \Sigma(a_1)^i\varepsilon_{t-i}$, and the known value of y_0. As an exercise, it is useful to show that iteration from y_t back to y_0 yields exactly the formula given by (1.18). Since $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$, it follows that

$$\begin{aligned} y_t &= a_0 + a_1[a_0 + a_1 y_{t-2} + \varepsilon_{t-1}] + \varepsilon_t \\ &= a_0(1+a_1) + a_1\varepsilon_{t-1} + \varepsilon_t + a_1^2[a_0 + a_1 y_{t-3} + \varepsilon_{t-2}] \end{aligned}$$

Continuing the iteration back to period 0 yields equation (1.18).

Iteration without an Initial Condition

Suppose you were not given the initial condition for y_0. The solution given by (1.18) would no longer be appropriate because the value of y_0 is an unknown. You could not select this initial value of y and iterate forward, nor could you iterate backward from y_t and simply choose to stop at $t = t_0$. Thus, suppose we continued to iterate backward by substituting $a_0 + a_1 y_{-1} + \varepsilon_0$ for y_0 in (1.18):

$$\begin{aligned} y_t &= a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t[a_0 + a_1 y_{-1} + \varepsilon_0] + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \\ &= a_0 \sum_{i=0}^{t} a_1^i + \sum_{i=0}^{t} a_1^i \varepsilon_{t-i} + a_1^{t+1} y_{-1} \end{aligned} \qquad (1.19)$$

Continuing to iterate backward another m periods, we obtain

$$y_t = a_0 \sum_{i=0}^{t+m} a_1^i + \sum_{i=0}^{t+m} a_1^i \varepsilon_{t-i} + a_1^{t+m+1} y_{-m-1} \qquad (1.20)$$

Now examine the pattern emerging from (1.19) and (1.20). If $|a_1| < 1$, the term a_1^{t+m+1} approaches zero as m approaches infinity. Also, the infinite sum $[1 + a_1 + (a_1)^2 + \ldots]$ converges to $1/(1 - a_1)$. Thus, if we temporarily assume that $|a_1| < 1$, after continual substitution, (1.20) can be written as

$$y_t = a_0/(1-a_1) + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \qquad (1.21)$$

You should take a few minutes to convince yourself that (1.21) is a solution to the original difference equation (1.17); substitution of (1.21) into (1.17) yields an

identity. However, (1.21) is not a unique solution. For any arbitrary value of A, a solution to (1.17) is given by

$$y_t = A a_1^t + a_0/(1-a_1) + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \qquad (1.22)$$

To verify that *for any arbitrary value of A* (1.22) is a solution, substitute (1.22) into (1.17) to obtain

$$a_0/(1-a_1) + Aa_1^t + \sum_{i=0}^{\infty} a_1^i\varepsilon_{t-i} = a_0 + a_1[a_0/(1-a_1) + Aa_1^{t-1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-1-i}] + \varepsilon_t$$

Since the two sides are identical, (1.22) is necessarily a solution to (1.17).

Reconciling the Two Iterative Methods

Given the iterative solution (1.22), suppose that you are now given an initial condition concerning the value of y in the arbitrary period t_0. It is straightforward to show that we can impose the initial condition on (1.22) to yield the same solution as (1.18). Since (1.22) must be valid for all periods (including t_0), when $t = 0$, it must be true that

$$y_0 = A + a_0/(1-a_1) + \sum_{i=0}^{\infty} a_1^i\varepsilon_{-i} \quad \text{so that}$$

$$A = y_0 - a_0/(1-a_1) - \sum_{i=0}^{\infty} a_1^i\varepsilon_{-i} \qquad (1.23)$$

Since y_0 is given, we can view (1.23) as the value of A that renders (1.22) a solution to (1.17) given the initial condition. Hence, the presence of the initial condition eliminates the arbitrariness of A. Substituting this value of A into (1.22) yields

$$y_t = [y_0 - a_0/(1-a_1) - \sum_{i=0}^{\infty} a_1^i\varepsilon_{-i}]a_1^t + a_0/(1-a_1) + \sum_{i=0}^{\infty} a_1^i\varepsilon_{t-i} \qquad (1.24)$$

Simplification of (1.24) results in

$$y_t = [y_0 - a_0/(1-a_1)]a_1^t + a_0/(1-a_1) + \sum_{i=0}^{t-1} a_1^i\varepsilon_{t-i} \qquad (1.25)$$

You should take a moment to verify that (1.25) is identical to (1.18).

Nonconvergent Sequences

Given that $|a_1| < 1$, (1.21) is the limiting value of (1.20) as m grows infinitely large. What happens to the solution in other circumstances? If $|a_1| > 1$, it is not possible to move from (1.20) to (1.21) because the expression $|a_1|^{t+m}$ grows infinitely large as $t+m$ approaches infinity.[1] However, if there is an initial condition, there is no need to obtain the infinite summation. Simply select the initial condition y_0 and iterate forward; the result will be (1.18):

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

Although the successive values of the $\{y_t\}$ sequence will become progressively larger in absolute value, all values in the series will be finite.

A very interesting case arises if $a_1 = 1$. Rewrite (1.17) as:

$$y_t = a_0 + y_{t-1} + \varepsilon_t$$

or

$$\Delta y_t = a_0 + \varepsilon_t$$

As you should verify by iterating from y_t back to y_0, a solution to this equation is[2]

$$y_t = a_0 t + \sum_{i=1}^{t} \varepsilon_i + y_0 \qquad (1.26)$$

After a moment's reflection, the form of the solution is quite intuitive. In every period t, the value of y_t changes by $a_0 + \varepsilon_t$ units. After t periods, there are t such changes; hence, the total change is ta_0 plus the t values of the $\{\varepsilon_t\}$ sequence. Notice that the solution contains summation of all disturbances from ε_1 through ε_t. Thus, when $a_1 = 1$, each disturbance has a permanent non-decaying effect on the value of y_t. You should compare this result to the solution found in (1.21). For the case in which $|a_1| < 1$, $|a_1|^t$ is a decreasing function of t so that the effects of past disturbances become successively smaller over time.

The importance of the magnitude of a_1 is illustrated in Figure 1.2. Twenty-five random numbers with a theoretical mean equal to zero were computer-generated and denoted by ε_1 through ε_{25}. Then the value of y_0 was set equal to unity and the next twenty-five values of the $\{y_t\}$ sequence were constructed using the formula $y_t = 0.9y_{t-1} + \varepsilon_t$. The result is shown by the thin line in Panel (*a*) of Figure 1.2. If you substitute $a_0 = 0$ and $a_1 = 0.9$ into (1.18), you will see that the time path of $\{y_t\}$ consists of two parts. The first part, 0.9^t, is shown by the slowly decaying thick line in the panel. This term dominates the solution for relatively small values of t. The influence of the random part is shown by the difference between the thin and the thick line; you can see that the first several values of $\{\varepsilon_t\}$ are negative. As t increases, the influence of the random component becomes more pronounced.

Using the previously drawn random numbers, we again set y_0 equal to unity and a second sequence was constructed using the formula $y_t = 0.5y_{t-1} + \varepsilon_t$. This second sequence is shown by the thin line in Panel (*b*) of Figure 1.2. The influence of the expression 0.5^t is shown by the rapidly decaying thick line. Again, as t increases, the random portion of the solution becomes more dominant in the time path of $\{y_t\}$. When we compare the first two panels, it is clear that reducing the magnitude of $|a_1|$ increases the rate of convergence. Moreover, the discrepancies between the simulated values of y_t and the thick line are less pronounced in the second panel. As you can see in (1.18), each value of ε_{t-i} enters the solution for y_t with a coefficient of $(a_1)^i$. The

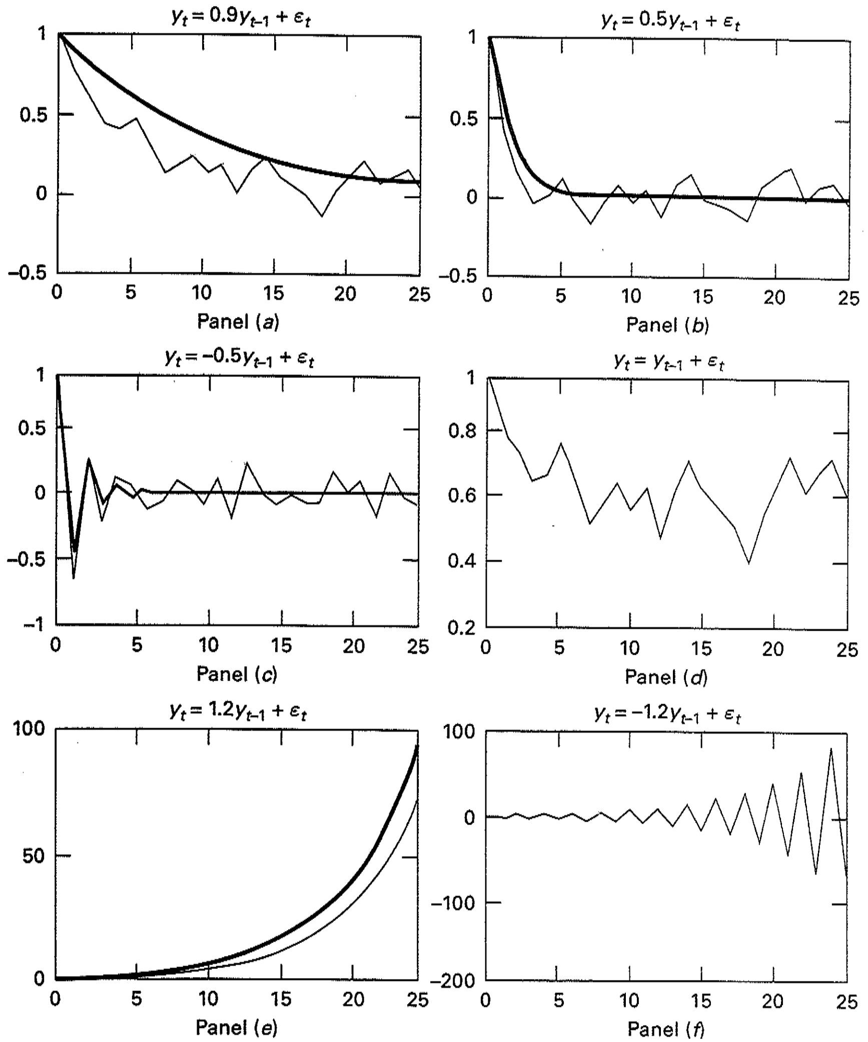


FIGURE 1.2 Convergent and Nonconvergent Sequences

smaller value of a_1 means that the past realizations of ε_{t-i} have a smaller influence on the current value of y_t.

Simulating a third sequence with $a_1 = -0.5$ yields the thin line shown in Panel (*c*). The oscillations are due to the negative value of a_1. The expression $(-0.5)^t$, shown by the thick line, is positive when t is even and negative when t is odd. Since $|a_1| < 1$, the oscillations are dampened.

The next three panels in Figure 1.2 all show nonconvergent sequences. Each uses the initial condition $y_0 = 1$ and the same twenty-five values of $\{\varepsilon_t\}$ used in the other simulations. The thin line in Panel (*d*) shows the time path of $y_t = y_{t-1} + \varepsilon_t$. Since each value of ε_t has an expected value of zero, Panel (*d*) illustrates a random-walk process. Here $\Delta y_t = \varepsilon_t$ so that the change in y_t is random. The nonconvergence is shown by the tendency of $\{y_t\}$ to meander. In Panel (*e*), the thick line representing the explosive expression $(1.2)^t$ dominates the random portion of the $\{y_t\}$ sequence. Also notice that the discrepancy between the simulated $\{y_t\}$ sequence and the thick line widens as t increases. The reason is that past values of ε_{t-i} enter the solution for y_t with the coefficient $(1.2)^i$. As i increases, the importance of these previous discrepancies becomes increasingly important. Similarly, setting $a_1 = -1.2$ results in the exploding oscillations shown in the lower-right panel of the figure. The value $(-1.2)^t$ is positive for even values of t and negative for odd values of t.

4. AN ALTERNATIVE SOLUTION METHODOLOGY

Solution by the iterative method breaks down in higher-order equations. The algebraic complexity quickly overwhelms any reasonable attempt to find a solution. Fortunately, there are several alternative solution techniques that can be helpful in solving the *n*th-order equation given by (1.10). If we use the principle that you should learn to walk before you learn to run, it is best to step through the first-order equation given by (1.17). Although you will be covering some familiar ground, the first-order case illustrates the general methodology extremely well. To split the procedure into its component parts, consider only the homogeneous portion of (1.17):[3]

$$y_t = a_1 y_{t-1} \tag{1.27}$$

The solution to this homogeneous equation is called the **homogeneous solution;** at times it will be useful to denote the homogeneous solution by the expression y_t^h. Obviously, the trivial solution $y_t = y_{t-1} = \ldots = 0$ satisfies (1.27). However, this solution is not unique. By setting a_0 and all values of $\{\varepsilon_t\}$ equal to zero, (1.18) becomes $y_t = a_1^t y_0$. Hence, $y_t = a_1^t y_0$ must be a solution to (1.27). Yet, even this solution does not constitute the full set of solutions. It is easy to verify that the expression a_1^t multiplied by any arbitrary constant A satisfies (1.27). Simply substitute $y_t = Aa_1^t$ and $y_{t-1} = Aa_1^{t-1}$ into (1.27) to obtain

$$Aa_1^t = a_1 A a_1^{t-1}$$

Since $a_1^t = a_1 a_1^{t-1}$, it follows that $y_t = Aa_1^t$ also solves (1.27). With the aid of the thick lines in Figure 1.2, we can classify the properties of the homogeneous solution as follows:

1. If $|a_1| < 1$, the expression a_1^t converges to zero as t approaches infinity. Convergence is direct if $0 < a_1 < 1$ and oscillatory if $-1 < a_1 < 0$.
2. If $|a_1| > 1$, the homogeneous solution is not stable. If $a_1 > 1$, the homogeneous solution approaches infinity as t increases. If $a_1 < -1$, the homogeneous solution oscillates explosively.
3. If $a_1 = 1$, any arbitrary constant A satisfies the homogeneous equation $y_t = y_{t-1}$. If $a_1 = -1$, the system is *meta-stable*: $a_1^t = 1$ for even values of t and -1 for odd values of t.

Now consider (1.17) in its entirety. In the last section, you confirmed that (1.21) is a valid solution to (1.17). Equation (1.21) is called a **particular solution** to the difference equation; all such particular solutions will be denoted by the term y_t^p. The term "particular" stems from the fact that a solution to a difference equation may not be unique; hence, (1.21) is just one particular solution out of the many possibilities.

In moving to (1.22) you verified that the particular solution was not unique. The homogeneous solution Aa_1^t plus the particular solution given by (1.21) constituted the complete solution to (1.17). The **general solution** to a difference equation is defined to be a particular solution plus all homogeneous solutions. Once the general solution is obtained, the arbitrary constant A can be eliminated by imposing an initial condition for y_0.

The Solution Methodology

The results of the first-order case are directly applicable to the *n*th-order equation given by (1.10). In this general case, it will be more difficult to find the particular solution and there will be n distinct homogeneous solutions. Nevertheless, the solution methodology will always entail the following four steps:

STEP 1: Form the homogeneous equation and find all n homogeneous solutions;
STEP 2: Find a particular solution;
STEP 3: Obtain the general solution as the sum of the particular solution and a linear combination of all homogeneous solutions;
STEP 4: Eliminate the arbitrary constant(s) by imposing the initial condition(s) on the general solution.

Before we address the various techniques that can be used to obtain homogeneous and particular solutions, it is worthwhile to illustrate the methodology using the equation:

$$y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3 \tag{1.28}$$

Clearly, this second-order equation is in the form of (1.10) with $a_0 = 3$, $a_1 = 0.9$, $a_2 = -0.2$, and $x_t = 0$. Beginning with the first of the four steps, form the homogeneous equation:

$$y_t - 0.9y_{t-1} + 0.2y_{t-2} = 0 \tag{1.29}$$

In the first-order case of (1.17), the homogeneous solution was Aa_1^t. Section 6 will show you how to find the complete set of homogeneous solutions. For now, it is sufficient to assert that the two homogeneous solutions are: $y_{1t}^h = (0.5)^t$ and $y_{2t}^h = (0.4)^t$. To verify the first solution, note that $y_{1t-1}^h = (0.5)^{t-1}$ and $y_{1t-2}^h = (0.5)^{t-2}$. Thus, y_{1t}^h is a solution if it satisfies

$$(0.5)^t - 0.9(0.5)^{t-1} + 0.2(0.5)^{t-2} = 0$$

If we divide by $(0.5)^{t-2}$, the issue is whether

$$(0.5)^2 - 0.9(0.5) + 0.2 = 0$$

Carrying out the algebra, $0.25 - 0.45 + 0.2$ does equal zero so that $(0.5)^t$ is a solution to (1.29). In the same way, it is easy to verify that $y_{2t}^h = (0.4)^t$ is a solution since

$$(0.4)^t - 0.9(0.4)^{t-1} + 0.2(0.4)^{t-2} = 0$$

Divide by $(0.4)^{t-2}$ to obtain $(0.4)^2 - 0.9(0.4) + 0.2 = 0.16 - 0.36 + 0.2 = 0$.

The second step is to obtain a particular solution; you can easily confirm that the particular solution $y_t^p = 10$ solves (1.28) as: $10 = 0.9(10) - 0.2(10) + 3$.

The third step is to combine the particular solution and a linear combination of both homogeneous solutions to obtain

$$y_t = A_1(0.5)^t + A_2(0.4)^t + 10$$

where A_1 and A_2 are arbitrary constants.

For the fourth step, assume you have two initial conditions for the $\{y_t\}$ sequence. So that we can keep our numbers reasonably round, suppose that $y_0 = 13$ and $y_1 = 11.3$. Thus, for periods zero and one, our solution must satisfy

$$13 = A_1 + A_2 + 10$$
$$11.3 = A_1(0.5) + A_2(0.4) + 10.$$

Solving simultaneously for A_1 and A_2, you should find $A_1 = 1$ and $A_2 = 2$. Hence, the solution is

$$y_t = (0.5)^t + 2(0.4)^t + 10$$

Generalizing the Method

To show that this method is applicable to higher-order equations, consider the homogeneous part of (1.10):

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} \qquad (1.30)$$

As shown in Section 6, there are n homogeneous solutions that satisfy (1.30). For now, it is sufficient to demonstrate the following proposition: *If y_t^h is a homogeneous solution to (1.30), Ay_t^h is also a solution for any arbitrary constant A.* By assumption, y_t^h solves the homogeneous equation so that

$$y_t^h = \sum_{i=1}^{n} a_i y_{t-i}^h \qquad (1.31)$$

The expression Ay_t^h is also a solution if:

$$Ay_t^h = \sum_{i=1}^{n} a_i A y_{t-i}^h \qquad (1.32)$$

We know (1.32) is satisfied because dividing each term by A yields (1.31). Now suppose that there are two separate solutions to the homogeneous equation denoted by y_{1t}^h and y_{2t}^h. It is straightforward to show that for any two constants A_1 and A_2, the linear combination $A_1 y_{1t}^h + A_2 y_{2t}^h$ is also a solution to the homogeneous equation. If $A_1 y_{1t}^h + A_2 y_{2t}^h$ is a solution to (1.30), it must satisfy

$$A_1 y_{1t}^h + A_2 y_{2t}^h = a_1 (A_1 y_{1t-1}^h + A_2 y_{2t-1}^h)$$
$$+ a_2 (A_1 y_{1t-2}^h + A_2 y_{2t-2}^h) + \dots + a_n (A_1 y_{1t-n}^h + A_2 y_{2t-n}^h)$$

Regrouping terms, we want to know if

$$(A_1 y_{1t}^h - \sum_{i=1}^{n} A_1 a_i y_{1t-i}^h) + (A_2 y_{2t}^h - \sum_{i=1}^{n} A_2 a_i y_{2t-i}^h) = 0$$

Since $A_1 y_{1t}^h$ and $A_2 y_{2t}^h$ are separate solutions to (1.30), each of the expressions in brackets is zero. Hence, the linear combination is necessarily a solution to the homogeneous equation. This result easily generalizes to all n homogeneous solutions to an nth-order equation.

Finally, the use of Step 3 is appropriate since *the sum of any particular solution and any linear combination of all homogeneous solutions is also a solution.* To prove this proposition, substitute the sum of the particular and homogeneous solutions into (1.10) to obtain

$$y_t^p + y_t^h = a_0 + \sum_{i=1}^{n} a_i (y_{t-i}^p + y_{t-i}^h) + x_t \qquad (1.33)$$

Recombining the terms in (1.33), we want to know if

$$[y_t^p - a_0 - \sum_{i=1}^{n} a_i y_{t-i}^p - x_t] + [y_t^h - \sum_{i=1}^{n} a_i y_{t-i}^h] = 0 \qquad (1.34)$$

Since y_t^p solves (1.10), the expression in the first bracket of (1.34) is zero. Since y_t^h solves the homogeneous equation, the expression in the second bracket is zero. Thus, (1.34) is an identity; the sum of the homogeneous and particular solutions solves (1.10).

5. THE COBWEB MODEL

An interesting way to illustrate the methodology outlined in the previous section is to consider a stochastic version of the traditional **cobweb model.** Since the model was originally developed to explain the volatility in agricultural prices, let the market for a product—say wheat—be represented by

$$d_t = a - \gamma p_t \qquad \gamma > 0 \qquad (1.35)$$

$$s_t = b + \beta p_t^* + \varepsilon_t \qquad \beta > 0 \qquad (1.36)$$

$$s_t = d_t \qquad (1.37)$$

where: d_t = demand for wheat in period t
s_t = supply of wheat in t
p_t = market price of wheat in t
p_t^* = price that farmers expect to prevail at t
ε_t = a zero mean stochastic supply shock

and parameters a, b, γ, and β are all positive such that $a > b$.[4]

The nature of the model is such that consumers buy as much wheat as is desired at the market clearing price p_t. At planting time, farmers do not know the price prevailing at harvest time; they base their supply decision on the expected price (p_t^*). The actual quantity produced depends on the planned quantity $b + \beta p_t^*$ plus a random supply shock ε_t. Once the product is harvested, market equilibrium requires that the quantity supplied equal the quantity demanded. Unlike the actual market for wheat, the model does not allow for the possibility of storage. The essence of the cobweb model is that farmers form their expectations in a naive fashion; let farmers use last year's price as the expected market price

$$p_t^* = p_{t-1} \tag{1.38}$$

Point E in Figure 1.3 represents the long-run equilibrium price and quantity combination. Note that the equilibrium concept in this stochastic model differs from that of the traditional cobweb model. If the system is stable, successive prices will *tend* to converge to point E. However, the nature of the stochastic equilibrium is such that the ever-present supply shocks prevent the system from remaining at E. Nevertheless, it is useful to solve for the long-run price. If we set all values of the $\{\varepsilon_t\}$ sequence equal to zero, set $p_t = p_{t-1} = \ldots = p$, and equate supply and demand, the long-run equilibrium price is given by $p = (a - b)/(\gamma+\beta)$. Similarly, the equilibrium quantity (s) is given by $s = (a\beta + \gamma b)/(\gamma + \beta)$.

To understand the dynamics of the system, suppose that farmers in t plan to produce the equilibrium quantity s. However, let there be a negative supply shock such that the actual quantity produced turns out to be s_t. As shown by point 1 in Figure 1.3, consumers are willing to pay p_t for the quantity s_t; hence, market equilibrium in t occurs at point 1. Updating one period allows us to see the main result of the cobweb model. For simplicity, assume that all subsequent values of the supply shock are zero (i.e., $\varepsilon_{t+1} = \varepsilon_{t+2} = \ldots = 0$). At the beginning of period $t + 1$, farmers expect the price at harvest time to be the price of the previous period; thus $p_{t+1}^* = p_t$. Accordingly, they produce and market quantity s_{t+1} (see point 2 in the figure); consumers, however, are

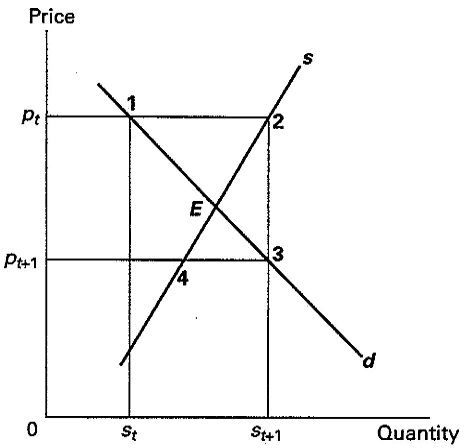


FIGURE 1.3 The Cobweb Model

willing to buy quantity s_{t+1} only if the price falls to that indicated by p_{t+1} (see point 3 in the figure). The next period begins with farmers expecting to be at point 4. The process continually repeats until the equilibrium point E is attained.

As drawn, Figure 1.3 suggests that the market will always converge to the long-run equilibrium point. This result does not hold for all demand and supply curves. To formally derive the stability condition, combine (1.35) through (1.38) to obtain

$$b + \beta p_{t-1} + \varepsilon_t = a - \gamma p_t$$

or

$$p_t = (-\beta/\gamma)p_{t-1} + (a - b)/\gamma - \varepsilon_t/\gamma \tag{1.39}$$

Clearly, (1.39) is a stochastic first-order linear difference equation with constant coefficients. To obtain the general solution, proceed using the four steps listed at the end of the last section:

1. Form the homogeneous equation: $p_t = (-\beta/\gamma)p_{t-1}$. In the next section you will learn how to find the solution(s) to a homogeneous equation. For now, it is sufficient to verify that the homogeneous solution is

$$p_t^h = A(-\beta/\gamma)^t$$

where A is an arbitrary constant.

2. If the ratio β/γ is less than unity, you can iterate (1.39) backward from p_t to verify that the particular solution for the price is

$$p_t^p = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} \tag{1.40}$$

If $\beta/\gamma \geq 1$, the infinite summation in (1.40) is not convergent. As discussed in the last section, it is necessary to impose an initial condition on (1.40) if $\beta/\gamma \geq 1$.

3. The general solution is the sum of the homogeneous and particular solutions; combining these two solutions, the general solution is

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} + A(-\beta/\gamma)^t \tag{1.41}$$

4. In (1.41), A is an arbitrary constant that can be eliminated if we know the price in some initial period. For convenience, let this initial period have a time subscript of zero. Since the solution must hold for every period, *including period zero,* it must be that case that

$$p_0 = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{-i} + A(-\beta/\gamma)^0$$

Since $(-\beta/\gamma)^0 = 1$, the value of A is given by

$$A = p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{-i}$$

Substituting this solution for A back into (1.41) yields

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} + \left(-\frac{\beta}{\gamma}\right)^t \left[p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_{i=0}^{\infty}(-\beta/\gamma)^i \varepsilon_{-i}\right]$$

and after simplification of the two summations

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{i=0}^{t-1}(-\beta/\gamma)^i \varepsilon_{t-i} + \left(-\frac{\beta}{\gamma}\right)^t \left[p_0 - \frac{a-b}{\gamma+\beta}\right] \qquad (1.42)$$

We can interpret (1.42) in terms of Figure 1.3. In order to focus on the stability of the system, temporarily assume that all values of the $\{\varepsilon_t\}$ sequence are zero. Subsequently, we will return to a consideration of the effects of supply shocks. If the system begins in long-run equilibrium, the initial condition is such that $p_0 = (a-b)/(\gamma+\beta)$. In this case, inspection of equation (1.42) indicates that $p_t = (a-b)/(\gamma+\beta)$. Thus, if we begin the process at point E, the system remains in long-run equilibrium. Instead, suppose that the process begins at a price below long-run equilibrium: $p_0 < (a-b)/(\gamma+\beta)$. Equation (1.42) tells us that p_1 is

$$p_1 = (a-b)/(\gamma+\beta) + [p_0 - (a-b)/(\gamma+\beta)]\,(-\beta/\gamma)^1 \qquad (1.43)$$

Since $p_0 < (a-b)/(\gamma+\beta)$ and $-\beta/\gamma < 0$, it follows that p_1 will be above the long-run equilibrium price $(a-b)/(\gamma+\beta)$. In period 2

$$p_2 = (a-b)/(\gamma+\beta) + [\,p_0 - (a-b)/(\gamma+\beta)]\,(-\beta/\gamma)^2$$

although $p_0 < (a-b)/(\gamma+\beta)$, $(-\beta/\gamma)^2$ is positive; hence, p_2 is below the long-run equilibrium. For the subsequent periods, note that $(-\beta/\gamma)^t$ will be positive for even values of t and negative for odd values of t. Just as we found graphically, the successive values of the $\{p_t\}$ sequence will oscillate above and below the long-run equilibrium price. Since $(\beta/\gamma)^t$ goes to zero if $\beta < \gamma$ and explodes if $\beta > \gamma$, the magnitude of β/γ determines whether the price actually converges toward the long-run equilibrium. If $\beta/\gamma < 1$, the oscillations will diminish in magnitude, and if $\beta/\gamma > 1$, the oscillations will be explosive.

The economic interpretation of this stability condition is straightforward. The slope of the supply curve (i.e., dp_t/ds_t) is $1/\beta$ and the absolute value of the slope of the demand curve [i.e., $-dp_t/d(d_t)$] is $1/\gamma$. If the supply curve is steeper than the demand curve, $1/\beta > 1/\gamma$ or $\beta/\gamma < 1$ so that the system is stable. This is precisely the case illustrated in Figure 1.3. As an exercise, you should draw a diagram with the demand curve steeper than the supply curve and show that the price oscillates and diverges from the long-run equilibrium.

Now consider the effects of the supply shocks. The contemporaneous effect of a supply shock on the price of wheat is the partial derivative of p_t with respect to ε_t; from (1.42):

$$\frac{\partial p_t}{\partial \varepsilon_t} = -\frac{1}{\gamma} \qquad (1.44)$$

Equation (1.44) is called the **impact multiplier** since it shows the impact effect of a change in ε_t on the price in t. In terms of Figure 1.3, a negative value of ε_t implies a price above the long-run price p; the price in t rises by $1/\gamma$ units for each unit decline in current period supply. Of course, this terminology is not specific to the cobweb model; in terms of the nth-order model given by (1.10), the impact multiplier is the partial derivative of y_t with respect to the partial change in the forcing process.[5]

The effects of the supply shock in t persist into future periods. Updating (1.42) by one period yields the **one-period multiplier:**

$$\frac{\partial p_{t+1}}{\partial \varepsilon_t} = -\frac{1}{\gamma}(-\beta/\gamma)$$
$$= \beta/\gamma^2$$

Point 3 in Figure 1.3 illustrates how the price in $t+1$ is affected by the negative supply shock in t. It is straightforward to derive the result that the effects of the supply shock decay over time. Since $\beta/\gamma < 1$, the absolute value of $\partial p_t/\partial\varepsilon_t$ exceeds $\partial p_{t+1}/\partial\varepsilon_t$. All of the multipliers can be derived analogously; updating (1.42) by two periods:

$$\partial p_{t+2}/\partial\varepsilon_t = -(1/\gamma)(-\beta/\gamma)^2$$

and after n periods:

$$\partial p_{t+n}/\partial\varepsilon_t = -(1/\gamma)(-\beta/\gamma)^n$$

The time path of all such multipliers is called the **impulse response function.** This function has many important applications in time-series analysis because it shows how the entire time path of a variable is affected by a stochastic shock. Here, the impulse function traces the effects of a supply shock in the wheat market. In other economic applications, you may be interested in the time path of a money supply shock or a productivity shock on real GDP.

In actuality, the function can be derived without updating (1.42) because it is always the case that

$$\frac{\partial p_{t+j}}{\partial \varepsilon_t} = \frac{\partial p_t}{\partial \varepsilon_{t-j}}$$

To find the impulse response function, simply find the partial derivative of (1.42) with respect to the various ε_{t-j}. These partial derivatives are nothing more than the coefficients of the $\{\varepsilon_{t-j}\}$ sequence in (1.42).

Each of the three components in (1.42) has a direct economic interpretation. The deterministic portion of the particular solution $(a-b)/(\gamma+\beta)$ is the long-run equilibrium price; if the stability condition is met, the $\{p_t\}$ sequence tends to converge to this long-run value. The stochastic component of the particular solution captures the short-run price adjustments due to the supply shocks. The ultimate decay of the coefficients of the impulse response function guarantees that the effects of changes in the various ε_t are of a short-run duration. The third component is the expression $(-\beta/\gamma)^t A = (-\beta/\gamma)^t[p_0 - (a-b)/(\gamma+\beta)]$. The value of A is the initial period's deviation of the price from its long-run equilibrium level. Given that $\beta/\gamma < 1$, the importance of this initial deviation diminishes over time.

6. SOLVING HOMOGENEOUS DIFFERENCE EQUATIONS

Higher-order difference equations arise quite naturally in economic analysis. Equation (1.5)—the reduced-form GDP equation resulting from Samuelson's (1939) model—is an example of a second-order difference equation. Moreover, in time-series econometrics it is quite typical to estimate second and higher-order equations. To begin our examination of homogeneous solutions, consider the second-order equation

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} = 0 \tag{1.45}$$

Given the findings in the first-order case, you should suspect that the homogeneous solution has the form $y_t^h = A\alpha^t$. Substitution of this trial solution into (1.45) yields

$$A\alpha^t - a_1 A\alpha^{t-1} - a_2 A\alpha^{t-2} = 0 \tag{1.46}$$

Clearly, any arbitrary value of A is satisfactory. If you divide (1.46) by $A\alpha^{t-2}$, the problem is to find the values of α that satisfy

$$\alpha^2 - a_1\alpha - a_2 = 0 \tag{1.47}$$

Solving this quadratic equation—called the **characteristic equation**—yields two values of α, called the **characteristic roots.** Using the quadratic formula, we find that the two characteristic roots are

$$\alpha_1, \alpha_2 = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2}$$

$$= (a_1 \pm \sqrt{d})/2 \tag{1.48}$$

where d is the discriminant $[(a_1)^2 + 4a_2]$.

Each of these two characteristic roots yields a valid solution for (1.45). Again, these solutions are not unique. In fact, for any two arbitrary constants A_1 and A_2, the linear combination $A_1(\alpha_1)^t + A_2(\alpha_2)^t$ also solves (1.45). As proof, simply substitute $y_t = A_1(\alpha_1)^t + A_2(\alpha_2)^t$ into (1.45) to obtain

$$A_1(\alpha_1)^t + A_2(\alpha_2)^t = a_1[A_1(\alpha_1)^{t-1} + A_2(\alpha_2)^{t-1}] + a_2[A_1(\alpha_1)^{t-2} + A_2(\alpha_2)^{t-2}]$$

Now, regroup terms as follows:

$$A_1[(\alpha_1)^t - a_1(\alpha_1)^{t-1} - a_2(\alpha_1)^{t-2}] + A_2[(\alpha_2)^t - a_1(\alpha_2)^{t-1} - a_2(\alpha_2)^{t-2}] = 0$$

Since α_1 and α_2 each solve (1.45), both terms in brackets must equal zero. As such, the complete homogeneous solution in the second-order case is

$$y_t^h = A_1(\alpha_1)^t + A_2(\alpha_2)^t$$

Without knowing the specific values of a_1 and a_2, we cannot find the two characteristic roots α_1 and α_2. Nevertheless, it is possible to characterize the nature of the solution; three possible cases are dependent on the value of the discriminant d.

CASE 1

If $a_1^2 + 4a_2 > 0$, d is a real number and there will be two distinct real characteristic roots. Hence, there are two separate solutions to the homogeneous equation denoted by $(\alpha_1)^t$ and $(\alpha_2)^t$. We already know that any linear combination of the two is also a solution. Hence,

$$y_t^h = A_1(\alpha_1)^t + A_2(\alpha_2)^t$$

It should be clear that if the absolute value of *either* α_1 or α_2 exceeds unity, the homogeneous solution will explode. Worksheet 1.1 examines two second-order equations showing real and distinct characteristic roots. In the first example, $y_t = 0.2y_{t-1} + 0.35y_{t-2}$, the characteristic roots are shown to be: $\alpha_1 = 0.7$ and $\alpha_2 = -0.5$. Hence, the full homogeneous solution is $y_t^h = A_1(0.7)^t + A_2(-0.5)^t$. Since both roots are less than unity in absolute value, the homogeneous solution is convergent. As you can see in the graph on the bottom left-hand side of Worksheet 1.1, convergence is not monotonic because of the influence of the expression $(-0.5)^t$.

In the second example, $y_t = 0.7y_{t-1} + 0.35y_{t-2}$. The worksheet indicates how to obtain the solution for the two characteristic roots. Given that one characteristic root is $(1.037)^t$, the $\{y_t\}$ sequence explodes. The influence of the negative root ($\alpha_2 = -0.337$) is responsible for the non-monotonicity of the time path. Since $(-0.337)^t$ quickly approaches zero, the dominant root is the explosive value 1.037.

CASE 2

If $a_1^2 + 4a_2 = 0$, it follows that $d = 0$ and $\alpha_1 = \alpha_2 = a_1/2$. Hence, a homogeneous solution is $a_1/2$. However, when $d = 0$, there is a second homogeneous solution given by $t(a_1/2)^t$. To demonstrate that $y_t^h = t(a_1/2)^t$ is a homogeneous solution, substitute it into (1.45) to determine whether

$$t(a_1/2)^t - a_1[(t-1)(a_1/2)^{t-1}] - a_2[(t-2)(a_1/2)^{t-2}] = 0$$

Divide by $(a_1/2)^{t-2}$ and form

$$-[(a_1^2/4) + a_2]t + [(a_1^2/2) + 2a_2] = 0$$

Since we are operating in the circumstance where $a_1^2 + 4a_2 = 0$, each bracketed expression is zero; hence, $t(a_1/2)^t$ solves (1.45). Again, for arbitrary constants A_1 and A_2, the complete homogeneous solution is

$$y_t^h = A_1(a_1/2)^t + A_2 t(a_1/2)^t$$

Clearly, the system is explosive if $|a_1| > 2$. If $|a_1| < 2$, the term $A_1(a_1/2)^t$ converges, but you might think that the effect of the term $t(a_1/2)^t$ is ambiguous [since the diminishing $(a_1/2)^t$ is multiplied by t]. This ambiguity is correct in the limited sense that the behavior of the homogeneous solution is not monotonic. As illustrated in Figure 1.4 for $a_1/2 = 0.95$, 0.9, and -0.9, as long as $|a_1| < 2$, lim[$t(a_1/2)^t$] is necessarily zero as $t \to \infty$; thus, there is always convergence. For $0 < a_1 < 2$, the homogeneous solution appears to explode before ultimately converging to zero. For $-2 < a_1 < 0$, the behavior is wildly erratic; the homogeneous solution appears to oscillate explosively before the oscillations dampen and finally converge to zero.

WORKSHEET 1.1

SECOND-ORDER EQUATIONS

Example 1: $y_t = 0.2y_{t-1} + 0.35y_{t-2}$. Hence: $a_1 = 0.2$ and $a_2 = 0.35$

Form the homogeneous equation: $y_t - 0.2y_{t-1} - 0.35y_{t-2} = 0$

A check of the discriminant reveals: $d = a_1^2 + 4a_2$ so that $d = 1.44$. Given that $d > 0$, the roots will be real and distinct.

Let the trial solution have the form: $y_t = \alpha^t$. Substitute the trial solution into the homogenous equation to obtain: $\alpha^t - 0.2\alpha^{t-1} - 0.35\alpha^{t-2} = 0$

Divide by α^{t-2} to obtain the characteristic equation: $\alpha^2 - 0.2\alpha^t - 0.35\alpha = 0$

Compute the two characteristic roots:

$\alpha_1 = 0.5(a_1 + d^{1/2})$ $\quad$ $\alpha_2 = 0.5(a_1 - d^{1/2})$

$\alpha_1 = 0.7$ $\quad$ $\alpha_2 = -0.5$

The homogeneous solution is: $A_1(0.7)^t + A_2(-0.5)^t$. The first graph shows the time path of this solution for the case in which the arbitrary constants equal unity and t runs from 1 to 20.

Example 2: $y_t = 0.7y_{t-1} + 0.35y_{t-2}$. Hence: $a_1 = 0.7$ and $a_2 = 0.35$

Form the homogeneous equation: $y_t - 0.7y_{t-1} - 0.35y_{t-2} = 0$

A check of the discriminant reveals: $d = a_1^2 + 4a_2$ so that $d = 1.89$. Given that $d > 0$, the roots will be real and distinct.

Form the characteristic equation $\alpha^t - 0.7\alpha^{t-1} - 0.35\alpha^{t-2} = 0$

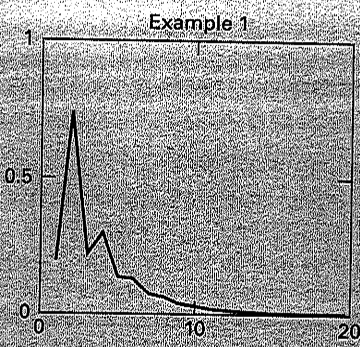
Divide by α^{t-2} to obtain the characteristic equation: $\alpha^2 - 0.7\alpha - 0.35\alpha = 0$

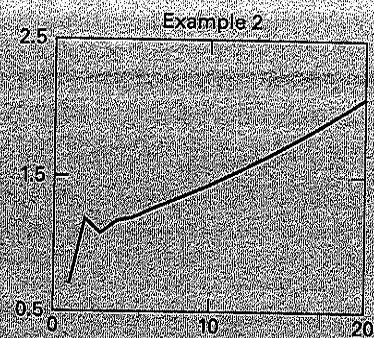
Compute the two characteristic roots:

$\alpha_1 = 0.5(a_1 + d^{1/2})$ $\quad$ $\alpha_2 = 0.5(a_1 - d^{1/2})$

$\alpha_1 = 1.037$ $\quad$ $\alpha_2 = -0.337$

The homogeneous solution is: $A_1(1.037)^t + A_2(-0.337)^t$. The second graph shows the time path of this solution for the case in which the arbitrary constants equal unity and t runs from 1 to 20.





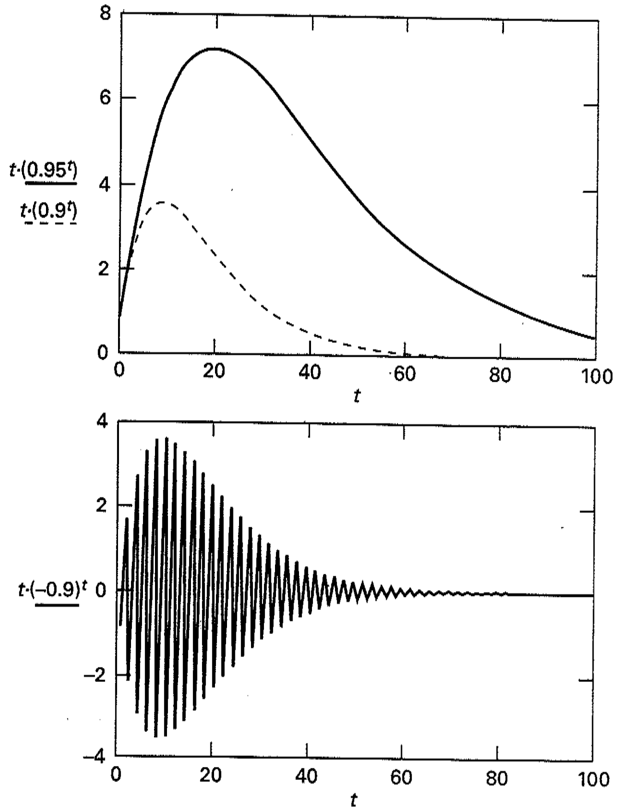


FIGURE 1-4 The Homogeneous Solution $t \cdot (a_1)^t$

CASE 3

If $a_1^2 + 4a_2 < 0$, it follows that d is negative so that the characteristic roots are imaginary. Since $a_1^2 \geq 0$, imaginary roots can occur only if $a_2 < 0$. Although this might be hard to interpret directly, if we switch to polar coordinates it is possible to transform the roots into more easily understood trigonometric functions. The technical details are presented in Appendix 1.1. For now, write the two characteristic roots as

$$\alpha_1 = (a_1 + i\sqrt{-d})/2 \qquad \alpha_2 = (a_1 - i\sqrt{-d})/2$$

where $i = \sqrt{-1}$.

As shown in Appendix 1.1, you can use de Moivre's theorem to write the homogeneous solution as

$$y_t^h = \beta_1 r^t \cos(\theta t + \beta_2) \tag{1.49}$$

where β_1 and β_2 are arbitrary constants, $r = (-a_2)^{1/2}$, and the value of θ is chosen so as to satisfy

$$\cos(\theta) = a_1/[2(-a_2)^{1/2}] \qquad (1.50)$$

The trigonometric functions impart a wavelike pattern to the time path of the homogeneous solution; note that the frequency of the oscillations is determined by θ. Since $\cos(\theta t) = \cos(2\pi + \theta t)$, the stability condition is determined solely by the magnitude of $r = (-a_2)^{1/2}$. If $|a_2| = 1$, the oscillations are of unchanging amplitude; the homogeneous solution is periodic. The oscillations will dampen if $|a_2| < 1$ and explode if $|a_2| > 1$.

Example: It is worthwhile to work through an exercise using an equation with imaginary roots. The left-hand side of Worksheet 1.2 examines the behavior of the equation $y_t = 1.6y_{t-1} - 0.9y_{t-2}$. A quick check shows that the discriminant d is negative so that the characteristic roots are imaginary. If we transform to polar coordinates, the value of r is given by $(0.9)^{1/2} = 0.949$. From (1.50), $\cos(\theta) = 1.6/(2*0.949) = 0.843$. You can use a trig table or a calculator to show that $\theta = 0.567$ (i.e., if $\cos(\theta) = 0.843$, $\theta = 0.567$). Thus, the homogeneous solution is

$$y_t^h = \beta_1(0.949)^t \cos(0.567t + \beta_2) \qquad (1.51)$$

The graph on the left-hand side of Worksheet 1.2 sets $\beta_1 = 1$ and $\beta_2 = 0$ and plots the homogeneous solution for $t = 1, \ldots, 30$. Example 2 uses the same value of a_2 (hence, $r = 0.949$) but sets $a_1 = -0.6$. Again, the value of d is negative; however, for this set of calculations, $\cos(\theta) = -0.316$ so that θ is 1.89. Comparing the two graphs, you can see that increasing the value of θ acts to increase the frequency of the oscillations.

Stability Conditions

The general stability conditions can be summarized using triangle ABC in Figure 1.5. Arc $A0B$ is the boundary between Cases 1 and 3; it is the locus of points where $d = a_1^2 + 4a_2 = 0$. The region above $A0B$ corresponds to Case 1 (since $d > 0$), and the region below $A0B$ corresponds to Case 3 (since $d < 0$).

In Case 1 (in which the roots are real and distinct), stability requires that the largest root be less than unity and the smallest root be greater than −1. The largest characteristic root, $\alpha_1 = (a_1 + \sqrt{d})/2$, will be less than unity if

$$a_1 + (a_1^2 + 4a_2)^{1/2} < 2 \quad \text{or} \quad (a_1^2 + 4a_2)^{1/2} < 2 - a_1$$

Hence, $a_1^2 + 4a_2 < 4 - 4a_1 + a_1^2$

or

$$a_1 + a_2 < 1 \qquad (1.52)$$

The smallest root, $\alpha_2 = (a_1 - \sqrt{d})/2$, will be greater than minus one if

$$a_1 - (a_1^2 + 4a_2)^{1/2} > -2 \quad \text{or} \quad 2 + a_1 > (a_1^2 + 4a_2)^{1/2}$$

WORKSHEET 1.2

IMAGINARY ROOTS

Example 1	Example 2
$y_t - 1.6y_{t-1} + 0.9y_{t-2}$	$y_t + 0.6y_{t-1} + 0.9y_{t-2}$

a) Check the discriminant $d = (a_1)^2 + 4a_2$

Example 1	Example 2
$d = (1.6)^2 + 4(-0.9)$ $= -1.04$	$d = (-0.6)^2 + 4(-0.9)$ $= -3.24$

Hence, the roots are imaginary. The homogeneous solution has the form

$$y_t^h = \beta_1 r^t \cos(\theta t + \beta_2)$$

where β_1 and β_2 are arbitrary constants.

b) Obtain the value of $r = (-a_2)^{1/2}$

Example 1	Example 2
$r = (0.9)^{1/2}$ $= 0.949$	$r = (0.9)^{1/2}$ $= 0.949$

c) Obtain θ from $\cos(\theta) = a_1/[2(-a_1)^{1/2}]$

Example 1	Example 2
$\cos(\theta) = 1.6/[2(0.9)^{1/2}]$ $= 0.843$	$\cos(\theta) = -0.6/[2(0.9)^{1/2}]$ $= -0.316$

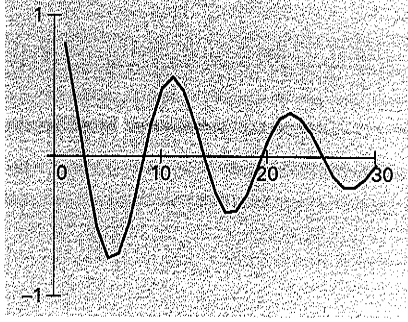
Given $\cos(\theta)$, use a trig-table to find θ

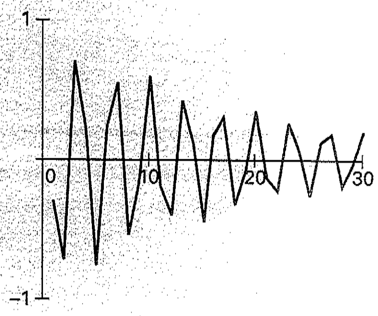
Example 1	Example 2
$\theta = 0.567$	$\theta = 1.89$

d) Form the homogeneous solution: $y_t^h = \beta_1 r^t \cos(\theta t + \beta_2)$

Example 1	Example 2
$y_t^h = \beta_1(0.949)^t \cos(0.567t + \beta_2)$	$y_t^h = \beta_1(0.949)^t \cos(1.89t + \beta_2)$

For $\beta_1 = 1$ and $\beta_2 = 0$, the time paths of the homogeneous solutions are





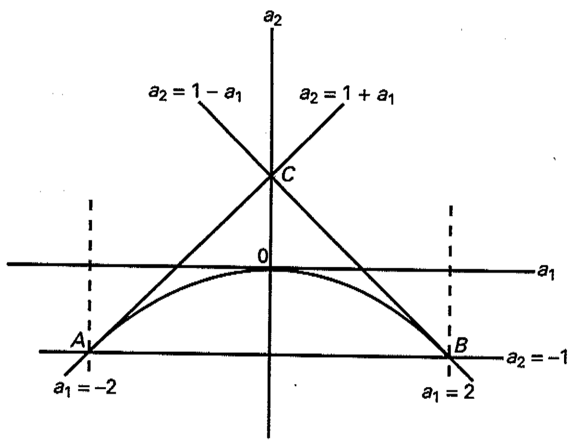


FIGURE 1.5 Characterizing the Stability Conditions

Hence: $4 + 4a_1 + a_1^2 > a_1^2 + 4a_2$

or

$$a_2 < 1 + a_1 \tag{1.53}$$

Thus, the region of stability in Case 1 consists of all points in the region bounded by $A0BC$. For any point in $A0BC$, conditions (1.52) and (1.53) hold and $d > 0$.

In Case 2 (repeated roots), $a_1^2 + 4a_2 = 0$. The stability condition is $|a_1| < 2$. Thus, the region of stability in Case 2 consists of all points on arc $A0B$. In Case 3 ($d < 0$), the stability condition is $r = (-a_2)^{1/2} < 1$. Hence

$$-a_2 < 1 \quad \text{(where } a_2 < 0\text{)} \tag{1.54}$$

Thus, the region of stability in Case 3 consists of all points in region $A0B$. For any point in $A0B$, (1.54) is satisfied and $d < 0$.

A succinct way to characterize the stability conditions is to state that the characteristic roots must lie within the unit circle. Consider the semicircle drawn in Figure 1.6. Real numbers are measured on the horizontal axis and imaginary numbers are measured on the vertical axis. If the characteristic roots α_1 and α_2 are both real, they can be plotted on the horizontal axis. Stability requires that they lie within a circle of radius one. Complex roots will lie somewhere in the complex plane. If $a_1 > 0$, the roots $\alpha_1 = (a_1 + i\sqrt{d})/2$ and $\alpha_2 = (a_1 - i\sqrt{d})/2$ can be represented by the two points shown in Figure 1.6. For example, α_1 is drawn by moving $a_1/2$ units along the real axis and $\sqrt{d}/2$ units along the imaginary axis. Using the distance formula, the length of the radius r is given by

$$r = \sqrt{(a_1/2)^2 + (d^{1/2}i/2)^2}$$

and, using the fact that $i^2 = -1$, we obtain

$$r = (-a_2)^{-1/2}$$

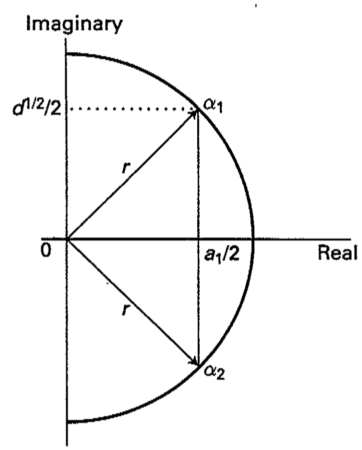


FIGURE 1.6 Characteristic Roots and the Unit Circle

The stability condition requires that $r < 1$. Therefore, when plotted on the complex plane, the two roots α_1 and α_2 must lie within a circle of radius equal to unity. In the time-series literature it is simply stated that *stability requires that all characteristic roots lie within the unit circle.*

Higher-Order Systems

The same method can be used to find the homogeneous solution to higher-order difference equations. The homogeneous equation for (1.10) is

$$y_t - \sum_{i=1}^{n} a_i y_{t-i} = 0 \tag{1.55}$$

Given the results in Section 4, you should suspect each homogeneous solution to have the form $y_t^h = A\alpha^t$ where A is an arbitrary constant. Thus, to find the value(s) of α, we seek the solution for

$$A\alpha^t - \sum_{i=1}^{n} a_i A\alpha^{t-i} = 0 \tag{1.56}$$

or, dividing through by α^{t-n}, we seek the values of α that solve

$$\alpha^n - a_1\alpha^{n-1} - a_2\alpha^{n-2} \ldots - a_n = 0 \tag{1.57}$$

This nth-order polynomial will yield n solutions for α. Denote these n characteristic roots by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Given the results in Section 4, the linear combination $A_1\alpha_1^t + A_2\alpha_2^t + \ldots + A_n\alpha_n^t$ is also a solution. The arbitrary constants A_1 through A_n can be eliminated by imposing n initial conditions on the general solution. The α_i may be real or complex numbers. Stability requires that all real valued α_i be less than unity in absolute value. Complex roots will necessarily come in pairs. Stability requires that all roots lie within the unit circle shown in Figure 1.6.

In most circumstances there is little need to directly calculate the characteristic roots of higher-order systems. Many of the technical details are included in Appendix 1.2 to this chapter. However, there are some useful rules for checking the stability conditions in higher-order systems.

1. In an nth-order equation, a necessary condition for all characteristic roots to lie inside the unit circle is

$$\sum_{i=1}^{n} a_i < 1$$

2. Since the values of the a_i can be positive or negative, a sufficient condition for all characteristic roots to lie inside the unit circle is

$$\sum_{i=1}^{n} |a_i| < 1$$

3. At least one characteristic root equals unity if

$$\sum_{i=1}^{n} a_i = 1$$

Any sequence that contains one or more characteristic roots that equal unity is called a **unit root** process.

4. For a third-order equation, the stability conditions can be written as

$$1 - a_1 - a_2 - a_3 > 0$$
$$1 + a_1 - a_2 + a_3 > 0$$
$$1 - a_1a_3 + a_2 - a_3^2 > 0$$
$$3 + a_1 + a_2 - 3a_3 > 0 \text{ or } 3 - a_1 + a_2 + 3a_3 > 0$$

Given that the first three inequalities are satisfied, either of the last two can be checked. One of the last conditions is redundant, given that the other three hold.

7. PARTICULAR SOLUTIONS FOR DETERMINISTIC PROCESSES

Finding the particular solution to a difference equation is often a matter of ingenuity and perseverance. The appropriate technique depends heavily on the form of the $\{x_t\}$ process. We begin by considering those processes that contain only deterministic components. Of course, in econometric analysis, the forcing process will contain both deterministic and stochastic components.

CASE 1

$x_t = 0$. When all elements of the $\{x_t\}$ process are zero, the difference equation becomes

$$y_t = a_0 + a_1y_{t-1} + a_2y_{t-2} + \ldots + a_ny_{t-n} \tag{1.58}$$

Intuition suggests that an unchanging value of y (i.e., $y_t = y_{t-1} = \ldots = c$) should solve the equation. Substitute the trial solution $y_t = c$ into (1.58) to obtain

$$c = a_0 + a_1c + a_2c + \ldots + a_nc$$

so that

$$c = a_0/(1 - a_1 - a_2 - \ldots - a_n) \tag{1.59}$$

As long as $(1 - a_1 - a_2 - \ldots - a_n)$ does not equal zero, the value of c given by (1.59) is a solution to (1.58). Hence, the particular solution to (1.58) is given by $y_t^p = a_0/(1 - a_1 - a_2 - \ldots - a_n)$.

If $1 - a_1 - a_2 - \ldots - a_n = 0$, the value of c in (1.59) is undefined; it is necessary to try some other form for the solution. The key insight is that $\{y_t\}$ is a unit root process if $\Sigma a_i = 1$. Since $\{y_t\}$ is not convergent, it stands to reason that the constant solution does not work. Instead, recall equations (1.12) and (1.26); these solutions suggest that a linear time trend can appear in the solution of a unit root process. As such, try the solution: $y_t^p = ct$. For ct to be a solution it must be the case that

$$ct = a_0 + a_1c(t-1) + a_2c(t-2) + \ldots + a_nc(t-n)$$

or, combining like terms,

$$(1 - a_1 - a_2 - \ldots - a_n)ct = a_0 - c(a_1 + 2a_2 + 3a_3 + \ldots + na_n)$$

Since $1 - a_1 - a_2 - \ldots - a_n = 0$, select the value of c such that

$$c = a_0/(a_1 + 2a_2 + 3a_3 + \ldots + na_n)$$

For example, let

$$y_t = 2 + 0.75y_{t-1} + 0.25y_{t-2}$$

Here, $a_1 = 0.75$ and $a_2 = 0.25$; $\{y_t\}$ is a unit root process because $a_1 + a_2 = 1$. The particular solution has the form ct, where $c = 2/[0.75 + 2(0.25)] = 1.6$. In the event that the solution ct fails, sequentially try the solutions: $y_t^p = ct^2, ct^3, \ldots, ct^n$. For an nth-order equation, one of these solutions will always be the particular solution.

CASE 2

The Exponential Case. Let x_t have the exponential form $b(d)^{rt}$, where b, d, and r are constants. Since r has the natural interpretation as a growth rate, we would expect to encounter this type of forcing process case in a growth context. We illustrate the solution procedure using the first-order equation

$$y_t = a_0 + a_1y_{t-1} + bd^{rt} \tag{1.60}$$

To try to gain an intuitive feel for the form of the solution, notice that if $b = 0$, (1.60) is a special case of (1.58). Hence, you should expect a constant to appear in the particular solution. Moreover, the expression d^{rt} grows at the constant rate r. Thus, you might expect the particular solution to have the form $y_t^p = c_0 + c_1d^{rt}$, where c_0 and c_1 are constants. If this equation is actually a solution, you should be able to substitute it back into (1.60) and obtain an identity. Making the appropriate substitutions, we get

$$c_0 + c_1 d^{rt} = a_0 + a_1[c_0 + c_1 d^{r(t-1)}] + bd^{rt} \tag{1.61}$$

For this solution to work, it is necessary to select c_0 and c_1 such that

$$c_0 = a_0/(1 - a_1) \text{ and } c_1 = [bd^r]/(d^r - a_1)$$

Thus, a particular solution is

$$y_t^p = \frac{a_0}{1 - a_1} + \frac{bd^r}{d^r - a_1} d^{rt}$$

The nature of the solution is that y_t^p equals the constant $a_0/(1 - a_1)$ plus an expression that grows at the rate r. Note that for $|d^r| < 1$, the particular solution converges to $a_0/(1 - a_1)$.

If either $a_1 = 1$ or $a_1 = d^r$, use the trick suggested in Case 1. If $a_1 = 1$, try the solution $c_0 = ct$, and if $a_1 = d^r$, try the solution $c_1 = tb$. Use precisely the same methodology in higher-order systems.

CASE 3

Deterministic Time Trend. In this case, let the $\{x_t\}$ sequence be represented by the relationship $x_t = bt^d$ where b is a constant and d is a positive integer. Hence

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + bt^d \tag{1.62}$$

Since y_t depends on t^d, it follows that y_{t-1} depends on $(t-1)^d$, y_{t-2} depends on $(t-2)^d$, and so on. As such, the particular solution has the form $y_t^p = c_0 + c_1 t + c_2 t^2 + \ldots + c_d t^d$. To find the values of the c_i, substitute the particular solution into (1.62). Then select the value of each c_i that results in an identity. Although various values of d are possible, in economic applications it is common to see models incorporating a linear time trend ($d = 1$). For illustrative purposes, consider the second-order equation $y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + bt$. Posit the solution $y_t^p = c_0 + c_1 t$ where c_0 and c_1 are undetermined coefficients. Substituting this "challenge solution" into the second-order difference equation yields

$$c_0 + c_1 t = a_0 + a_1[c_0 + c_1(t - 1)] + a_2[c_0 + c_1(t - 2)] + bt \tag{1.63}$$

Now select values of c_0 and c_1 so as to force equation (1.63) to be an identity for all possible values of t. If we combine all constant terms and all terms involving t, the required values of c_0 and c_1 are

$$c_1 = b/(1 - a_1 - a_2)$$
$$c_0 = [a_0 - (2a_2 + a_1)c_1] / (1 - a_1 - a_2)$$

so that

$$c_0 = [a_0/(1 - a_1 - a_2)] - [b/(1 - a_1 - a_2)^2](2a_2 + a_1)$$

Thus, the particular solution will also contain a linear time trend. You should have no difficulty foreseeing the solution technique if $a_1 + a_2 = 1$. In this circumstance—which is applicable to higher order cases, as well—try multiplying the original challenge solution by t.

8. THE METHOD OF UNDETERMINED COEFFICIENTS

At this point, it is appropriate to introduce the first of two useful methods for finding particular solutions when there are stochastic components in the $\{y_t\}$ process. The key insight of the **method of undetermined coefficients** is that the particular solution to a linear difference equation is necessarily linear. Moreover, the solution can depend only on time, a constant, and the elements of the forcing process $\{x_t\}$. Thus, it is often possible to know the exact form of the solution even though the coefficients of the solution are unknown. The technique involves positing a solution—called a **challenge solution**—that is a linear function of all terms thought to appear in the actual solution. The problem becomes one of finding the set of values for those undetermined coefficients that solve the difference equation.

The actual technique for finding the coefficients is straightforward. Substitute the challenge solution into the original difference equation and solve for the values of the undetermined coefficients that yield an identity for all possible values of the included variables. If it is not possible to obtain an identity, the form of the challenge solution is incorrect. Try a new trial solution and repeat the process. In fact, we used the method of undetermined coefficients when positing the challenge solutions $y_t^p = c_0 + c_1 d^{rt}$ and $y_t^p = c_0 + c_1 t$ for Cases 2 and 3 in Section 7.

To begin, reconsider the simple first-order equation $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$. Since you have solved this equation using the iterative method, the equation is useful for illustrating the method of undetermined coefficients. The nature of the $\{y_t\}$ process is such that the particular solution can depend only on a constant term, time, and the individual elements of the $\{\varepsilon_t\}$ sequence. Since t does not explicitly appear in the forcing process, t can be in the particular solution only if the characteristic root is unity. Since the goal is to illustrate the method, posit the challenge solution:

$$y_t = b_0 + b_1 t + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} \tag{1.64}$$

where b_0, b_1, and all the α_i are the coefficients to be determined.

Substitute (1.64) into the original difference equation to form

$$b_0 + b_1 t + \alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \ldots$$
$$= a_0 + a_1[b_0 + b_1(t - 1) + \alpha_0 \varepsilon_{t-1} + \alpha_1 \varepsilon_{t-2} + \ldots] + \varepsilon_t$$

Collecting like terms, we obtain

$$(b_0 - a_0 - a_1 b_0 + a_1 b_1) + b_1(1 - a_1)t + (\alpha_0 - 1)\varepsilon_t$$
$$+ (\alpha_1 - a_1\alpha_0)\varepsilon_{t-1} + (\alpha_2 - a_1\alpha_1)\varepsilon_{t-2} + (\alpha_3 - a_1\alpha_2)\varepsilon_{t-3} + \ldots = 0 \tag{1.65}$$

Equation (1.65) must hold for all values of t and all possible values of the $\{\varepsilon_t\}$ sequence. Thus, each of the following conditions must hold:

$$\alpha_0 - 1 = 0$$
$$\alpha_1 - a_1\alpha_0 = 0$$
$$\alpha_2 - a_1\alpha_1 = 0$$
$$\vdots$$
$$b_0 - a_0 - a_1 b_0 + a_1 b_1 = 0$$
$$b_1 - a_1 b_1 = 0$$

Notice that the first set of conditions can be solved for the α_i recursively. The solution of the first condition entails setting $\alpha_0 = 1$. Given this solution for α_0, the next equation requires $\alpha_1 = a_1$. Moving down the list, $\alpha_2 = a_1\alpha_1$ or $\alpha_2 = a_1^2$. Continuing the recursive process, we find $\alpha_i = a_1^i$. Now consider the last two equations. There are two possible cases depending on the value of a_1. If $a_1 \neq 1$, it immediately follows that $b_1 = 0$ and $b_0 = a_0/(1 - a_1)$. For this case, the particular solution is

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

Compare this result to (1.21); you will see that it is precisely the same solution found using the iterative method. The general solution is the sum of this particular solution plus the homogeneous solution Aa_1^t. Hence, the general solution is

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + Aa_1^t$$

Now, if there is an initial condition for y_0, it follows that

$$y_0 = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{-i} + A$$

Combining these two equations so as to eliminate the arbitrary constant A, we obtain

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + a_1^t\left[y_0 - a_0/(1-a_1) - \sum_{i=0}^{\infty} a_1^i \varepsilon_{-i}\right]$$

so that

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} + a_1^t\left[y_0 - a_0/(1-a_1)\right] \qquad (1.66)$$

It can be easily verified that (1.66) is identical to (1.25). Instead, if $a_1 = 1$, b_0 can be any arbitrary constant and $b_1 = a_0$. The improper form of the solution is

$$y_t = b_0 + a_0 t + \sum_{i=0}^{\infty} \varepsilon_{t-i}$$

The form of the solution is "improper" because the sum of the $\{\varepsilon_t\}$ sequence may not be finite. Therefore, it is necessary to impose an initial condition. If the value y_0 is given, it follows that

$$y_0 = b_0 + \sum_{i=0}^{\infty} \varepsilon_{-i}$$

Imposing the initial condition on the improper form of the solution yields (1.26)

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

To take a second example, consider the equation

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} \qquad (1.67)$$

Again, the solution can depend only on a constant, the elements of the $\{\varepsilon_t\}$ sequence, and t raised to the first power. As in the previous example, t does not need to be included in the challenge solution if the characteristic root differs from unity. To reinforce this point, use the challenge solution given by (1.64). Substitute this tentative solution into (1.67) to obtain

$$b_0 + b_1 t + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} = a_0 + a_1\left[b_0 + b_1(t-1) + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-1-i}\right] + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

Matching coefficients on all terms containing $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots,$ yields

$$\alpha_0 = 1$$
$$\alpha_1 = a_1\alpha_0 + \beta_1 \qquad [\text{so that } \alpha_1 = a_1 + \beta_1]$$
$$\alpha_2 = a_1\alpha_1 \qquad [\text{so that } \alpha_2 = a_1(a_1 + \beta_1)]$$
$$\alpha_3 = a_1\alpha_2 \qquad [\text{so that } \alpha_3 = (a_1)^2(a_1 + \beta_1)]$$
$$\ldots$$
$$\alpha_i = a_1\alpha_{i-1} \qquad [\text{so that } \alpha_i = (a_1)^{i-1}(a_1 + \beta_1)]$$

Matching coefficients of intercept terms and coefficients of terms containing t, we get

$$b_0 = a_0 + a_1 b_0 - a_1 b_1$$
$$b_1 = a_1 b_1$$

Again, there are two cases. If $a_1 \neq 1$, then $b_1 = 0$ and $b_0 = a_0/(1 - a_1)$. The particular solution is

$$y_t = \frac{a_0}{1-a_1} + \varepsilon_t + (a_1 + \beta_1)\sum_{i=1}^{\infty} a_1^{i-1} \varepsilon_{t-i}$$

The general solution augments the particular solution with the term Aa_1^t. You are left with the exercise of imposing the initial condition for y_0 on the general solution. Now consider the case in which $a_1 = 1$. The undetermined coefficients are such that $b_1 = a_0$ and b_0 is an arbitrary constant. The improper form of the solution is

$$y_t = b_0 + a_0 t + \varepsilon_t + (1 + \beta_1)\sum_{i=1}^{\infty} \varepsilon_{t-i}$$

If y_0 is given, it follows that

$$y_0 = b_0 + \varepsilon_0 + (1+\beta_1)\sum_{i=1}^{\infty}\varepsilon_{-i}$$

Hence, imposing the initial condition, we obtain

$$y_t = y_0 + a_0 t + \varepsilon_t + (1+\beta_1)\sum_{i=1}^{t-1}\varepsilon_{t-i}$$

Higher-Order Systems

The identical procedure is used for higher-order systems. As an example, let us find the particular solution to the second-order equation

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \qquad (1.68)$$

Since we have a second-order equation, we use the challenge solution

$$y_t = b_0 + b_1 t + b_2 t^2 + \alpha_0\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \dots$$

where b_0, b_1, b_2, and the α_i are the undetermined coefficients.

Substituting the challenge solution into (1.68) yields

$$\begin{aligned}[b_0+b_1t+b_2t^2] + \alpha_0\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \dots &= a_0 + a_1[b_0 + b_1(t-1) + b_2(t-1)^2 \\ &+ \alpha_0\varepsilon_{t-1} + \alpha_1\varepsilon_{t-2} + \alpha_2\varepsilon_{t-3} + \dots] + a_2[b_0 + b_1(t-2) + b_2(t-2)^2 \\ &+ \alpha_0\varepsilon_{t-2} + \alpha_1\varepsilon_{t-3} + \alpha_2\varepsilon_{t-4} + \dots] + \varepsilon_t\end{aligned}$$

There are several necessary and sufficient conditions for the values of the α_i's to render the equation above an identity for all possible realizations of the $\{\varepsilon_t\}$ sequence:

$$\begin{aligned}\alpha_0 &= 1 \\ \alpha_1 &= a_1\alpha_0 && [\text{so that } \alpha_1 = a_1] \\ \alpha_2 &= a_1\alpha_1 + a_2\alpha_0 && [\text{so that } \alpha_2 = (a_1)^2 + a_2] \\ \alpha_3 &= a_1\alpha_2 + a_2\alpha_1 && [\text{so that } \alpha_3 = (a_1)^3 + 2a_1a_2] \\ &\cdot \\ &\cdot\end{aligned}$$

Notice that for any value of $j \geq 2$, the coefficients solve the second-order difference equation $\alpha_j = a_1\alpha_{j-1} + a_2\alpha_{j-2}$. Since we know α_0 and α_1, we can solve for all the α_j iteratively. The properties of the coefficients will be precisely those discussed when considering homogeneous solutions:

1. Convergence necessitates that $|a_2| < 1$, $a_1 + a_2 < 1$, and that $a_2 - a_1 < 1$. Notice that convergence implies that past values of the $\{\varepsilon_t\}$ sequence ultimately have a successively smaller influence on the current value of y_t.
2. If the coefficients converge, convergence will be direct or oscillatory if $(a_1^2 + 4a_2) > 0$, will follow a sine/cosine pattern if $(a_1^2 + 4a_2) < 0$, and will "explode" and then converge if $(a_1^2 + 4a_2) = 0$. Appropriately setting the α_i, we are left with the remaining expression:

$$\begin{aligned}&b_2(1 - a_1 - a_2)t^2 + [b_1(1 - a_1 - a_2) + 2b_2(a_1 + 2a_2)]t + \\ &[b_0(1 - a_1 - a_2) - a_0 + a_1(b_1 - b_2) + 2a_2(b_1 - 2b_2)] = 0\end{aligned} \qquad (1.69)$$

Equation (1.69) must equal zero for all values of t. First, consider the case in which $a_1 + a_2 \neq 1$. Since $(1 - a_1 - a_2)$ does not vanish, it is necessary to set the value of b_2 equal to zero. Given that $b_2 = 0$ and that the coefficient of t must equal zero, it follows that b_1 must also be set equal to zero. Finally, given that $b_1 = b_2 = 0$, we must set $b_0 = a_0/(1 - a_1 - a_2)$. Instead, if $a_1 + a_2 = 1$, the solutions for the b_i depend on the specific values of a_0, a_1, and a_2. The key point is that *the stability condition for the homogeneous equation is precisely the condition for convergence of the particular solution. If any characteristic root of the homogeneous equation is equal to unity, a polynomial time trend will appear in the particular solution. The order of the polynomial is the number of unitary characteristic roots.* This result generalizes to higher-order equations.

If you are really clever, you can combine the discussion of the last section with the method of undetermined coefficients. Find the deterministic portion of the particular solution using the techniques discussed in the last section. Then use the method of undetermined coefficients to find the stochastic portion of the particular solution. In (1.67), for example, set $\varepsilon_t = \varepsilon_{t-1} = 0$ and obtain the solution $a_0/(1 - a_1)$. Now use the method of undetermined coefficients to find the particular solution of $y_t = a_1y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$. Add the deterministic and stochastic components to obtain all components of the particular solution.

A Solved Problem

To illustrate the methodology using a second-order equation, augment (1.28) with the stochastic term ε_t so that

$$y_t = 3 + 0.9y_{t-1} - 0.2y_{t-2} + \varepsilon_t \qquad (1.70)$$

You have already verified that the two homogeneous solutions are $A_1(0.5)^t$ and $A_2(0.4)^t$ and that the deterministic portion of the particular solution is $y_t^p = 10$. To find the stochastic portion of the particular solution, form the challenge solution

$$y_t = \sum_{i=0}^{\infty}\alpha_i\varepsilon_{t-i}$$

In contrast to (1.64), the intercept term b_0 is excluded (since we have already found the deterministic portion of the particular solution) and the time trend b_1t is excluded (since both characteristic roots are less than unity). For this challenge to work, it must satisfy

$$\begin{aligned}\alpha_0\varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \alpha_3\varepsilon_{t-3} + \dots &= 0.9[\alpha_0\varepsilon_{t-1} + \alpha_1\varepsilon_{t-2} + \alpha_2\varepsilon_{t-3} + \alpha_3\varepsilon_{t-4} + \dots] \\ &- 0.2\,[\alpha_0\varepsilon_{t-2} + \alpha_1\varepsilon_{t-3} + \alpha_2\varepsilon_{t-4} + \alpha_3\varepsilon_{t-5} + \dots] + \varepsilon_t\end{aligned} \qquad (1.71)$$

Since (1.71) must hold for all possible realizations of ε_t, ε_{t-1}, ε_{t-2}, …, each of the following conditions must hold:

$$\begin{aligned}\alpha_0 &= 1 \\ \alpha_1 &= 0.9\alpha_0\end{aligned}$$

so that $\alpha_1 = 0.9$, and for all $i \geq 2$,

$$\alpha_i = 0.9\alpha_{i-1} - 0.2\alpha_{i-2} \qquad (1.72)$$

Now, it is possible to solve (1.72) iteratively so that $\alpha_2 = 0.9\alpha_1 - 0.2\alpha_0 = 0.61$, $\alpha_3 = 0.9(0.61) - 0.2(0.9) = 0.369$, and so forth. A more elegant solution method is to view (1.72) as a second-order difference equation in the $\{\alpha_i\}$ sequence with initial conditions $\alpha_0 = 1$ and $\alpha_1 = 0.9$. The solution to (1.72) is

$$\alpha_i = 5(0.5)^i - 4(0.4)^i \tag{1.73}$$

To obtain (1.73), note that the solution to (1.72) is: $\alpha_i = A_3(0.5)^i + A_4(0.4)^i$ where A_3 and A_4 are arbitrary constants. Imposing the conditions $\alpha_0 = 1$ and $\alpha_1 = 0.9$ yields (1.73). If we use (1.73), it follows that: $\alpha_0 = 5(0.5)^0 - 4(0.4)^0 = 1$; $\alpha_1 = 5(0.5)^1 - 4(0.4)^1 = 0.9$; $\alpha_2 = 5(0.5)^2 - 4(0.4)^2 = 0.61$; and so on.

The general solution to (1.70) is the sum of the two homogeneous solutions and the deterministic and stochastic portions of the particular solution:

$$y_t = 10 + A_1(0.5)^t + A_2(0.4)^t + \sum_{i=0}^{\infty}\alpha_i\varepsilon_{t-i} \tag{1.74}$$

where the α_i are given by (1.73).

Given initial conditions for y_0 and y_1, it follows that A_1 and A_2 must satisfy

$$y_0 = 10 + A_1 + A_2 + \sum_{i=0}^{\infty}\alpha_i\varepsilon_{-i} \tag{1.75}$$

$$y_1 = 10 + A_1(0.5) + A_2(0.4) + \sum_{i=0}^{\infty}\alpha_i\varepsilon_{1-i} \tag{1.76}$$

Although the algebra gets messy, (1.75) and (1.76) can be substituted into (1.74) to eliminate the arbitrary constants:

$$y_t = 10 + (0.4)^t[5(y_0 - 10) - 10(y_1 - 10)]$$
$$+ (0.5)^t[10(y_1 - 10) - 4(y_0 - 10)] + \sum_{i=0}^{t-2}\alpha_i\varepsilon_{t-i}$$

9. LAG OPERATORS

If it is not important to know the actual values of the coefficients appearing in the particular solution, it is often more convenient to use lag operators rather than the method of undetermined coefficients. The **lag operator** L is defined to be a *linear* operator such that for any value y_t

$$L^i y_t \equiv y_{t-i} \tag{1.77}$$

Thus, L^i preceding y_t simply means to lag y_t by i periods. It is useful to remember the following properties of lag operators:

1. The lag of a constant is a constant: $Lc = c$.
2. The distributive law holds for lag operators. We can set $(L^i + L^j)y_t = L^i y_t + L^j y_t = y_{t-i} + y_{t-j}$.

3. The associative law of multiplication holds for lag operators. We can set $L^iL^jy_t = L^i(L^jy_t) = L^iy_{t-j} = y_{t-i-j}$. Similarly, we can set $L^iL^jy_t = L^{i+j}y_t = y_{t-i-j}$. Note that $L^0y_t = y_t$.
4. L raised to a negative power is actually a *lead* operator: $L^{-i}y_t = y_{t+i}$. To explain, define $j = -i$ and form $L^jy_t = y_{t-j} = y_{t+i}$.
5. For $|a| < 1$, the infinite sum $(1 + aL + a^2L^2 + a^3L^3 + \ldots)y_t = y_t/(1 - aL)$. This property of lag operators may not seem intuitive, but it follows directly from properties 2 and 3 above.

   Proof: Multiply each side by $(1 - aL)$ to form $(1 - aL)(1 + aL + a^2L^2 + a^3L^3 + \ldots)y_t = y_t$. Multiply the two expressions to obtain $(1 - aL + aL - a^2L^2 + a^2L^2 - a^3L^3 + \ldots)y_t = y_t$. Given that $|a| < 1$, the expression $a^nL^ny_t$ converges to zero as $n \to \infty$. Thus, the two sides of the equation are equal.
6. For $|a| > 1$, the infinite sum $[1 + (aL)^{-1} + (aL)^{-2} + (aL)^{-3} + \ldots]y_t = -aLy_t/(1 - aL)$. Thus,

   $$y_t/(1 - aL) = -(aL)^{-1}\sum_{i=0}^{\infty}(aL)^{-i}y_t$$

   Proof: Multiply by $(1 - aL)$ to form $(1 - aL)[1 + (aL)^{-1} + (aL)^{-2} + (aL)^{-3} + \ldots]y_t = -aLy_t$. Perform the indicated multiplication to obtain $[1 - aL + (aL)^{-1} - 1 + (aL)^{-2} - (aL)^{-1} + (aL)^{-3} - (aL)^{-2} \ldots]y_t = -aLy_t$. Given that $|a| > 1$, the expression $a^{-n}L^{-n}y_t$ converges to zero as $n \to \infty$. Thus, the two sides of the equation are equal.

Lag operators provide a concise notation for writing difference equations. Using lag operators, we can write the pth-order equation $y_t = a_0 + a_1y_{t-1} + \ldots + a_py_{t-p} + \varepsilon_t$ as

$$(1 - a_1L - a_2L^2 - \ldots - a_pL^p)y_t = a_0 + \varepsilon_t$$

or, more compactly, as

$$A(L)y_t = a_0 + \varepsilon_t$$

where $A(L)$ is the polynomial $(1 - a_1L - a_2L^2 - \ldots - a_pL^p)$

Since $A(L)$ can be viewed as a polynomial in the lag operator, the notation $A(1)$ is used to denote the sum of the coefficients

$$A(1) = 1 - a_1 - a_2 \ldots - a_p$$

As a second example, lag operators can be used to express the equation $y_t = a_0 + a_1y_{t-1} + \ldots + a_py_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q}$ as

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

where $A(L)$ and $B(L)$ are polynomials of orders p and q, respectively.

It is straightforward to use lag operators to solve linear difference equations. Again consider the first-order equation $y_t = a_0 + a_1y_{t-1} + \varepsilon_t$ where $|a_1| < 1$. Use the definition of L to form

$$y_t = a_0 + a_1Ly_t + \varepsilon_t \tag{1.78}$$

Solving for y_t, we obtain

$$y_t = \frac{a_0 + \varepsilon_t}{1 - a_1 L} \tag{1.79}$$

From property 1, we know that $La_0 = a_0$, so that $a_0/(1 - a_1 L) = a_0 + a_1 a_0 + a_1^2 a_0 + \ldots = a_0/(1 - a_1)$. From property 5, we know that $\varepsilon_t/(1 - a_1 L) = \varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \ldots$. Combining these two parts of the solution, we obtain the particular solution given by (1.21).

For practice, we can use lag operators to solve (1.67): $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$, where $|a_1| < 1$. Use property 2 to form $(1 - a_1 L)y_t = a_0 + (1 + \beta_1 L)\varepsilon_t$. Solving for y_t yields

$$y_t = [a_0 + (1 + \beta_1 L)\varepsilon_t]/(1 - a_1 L)$$

so that

$$y_t = [a_0 / (1 - a_1)] + [\varepsilon_t / (1 - a_1 L)] + [\beta_1\varepsilon_{t-1} / (1 - a_1 L)] \tag{1.80}$$

Expanding the last two terms of (1.80) yields the same solution found using the method of undetermined coefficients.

Now suppose $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$ but $|a_1| > 1$. The application of property 5 to (1.79) is inappropriate because it implies that y_t is infinite. Instead, expand (1.79) using property 6:

$$y_t = \frac{a_0}{1 - a_1} - (a_1 L)^{-1} \sum_{i=0}^{\infty} (a_1 L)^{-i} \varepsilon_t$$

$$= \frac{a_0}{1 - a_1} - \frac{1}{a_1} \sum_{i=0}^{\infty} (a_1 L)^{-i} \varepsilon_{t+1} \tag{1.81}$$

$$= \frac{a_0}{1 - a_1} - \left(\frac{1}{a_1}\right) \sum_{i=0}^{\infty} a_1^{-i} \varepsilon_{t+1+i} \tag{1.82}$$

Lag Operators in Higher-Order Systems

We can also use lag operators to transform the nth-order equation $y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_n y_{t-n} + \varepsilon_t$ into

$$(1 - a_1 L - a_2 L^2 - \ldots - a_n L^n) y_t = a_0 + \varepsilon_t$$

or

$$y_t = (a_0 + \varepsilon_t)/(1 - a_1 L - a_2 L^2 - \ldots a_n L^n)$$

From our previous analysis (also see Appendix 1.2), we know that the stability condition is such that the characteristic roots of the equation $\alpha^n - a_1\alpha^{n-1} - \ldots - a_n = 0$ all lie *within* the unit circle. Notice that the values of α solving the characteristic equation are the reciprocals of the values of L that solve the equation $1 - a_1 L \ldots - a_n L^n = 0$. In fact, the expression $1 - a_1 L \ldots - a_n L^n$ is often called the *inverse characteristic equation*. Thus, in the literature, it is often stated that the stability condition is for the characteristic roots of $(1 - a_1 L \ldots - a_n L^n)$ to lie *outside* of the unit circle.

In principle, one could use lag operators to actually obtain the coefficients of the particular solution. To illustrate using the second-order case, consider $y_t = (a_0 + \varepsilon_t)/(1 - a_1 L - a_2 L^2)$. If we knew the factors of the quadratic equation were such that $(1 - a_1 L - a_2 L^2) = (1 - b_1 L)(1 - b_2 L)$, we could write

$$y_t = (a_0 + \varepsilon_t)/[(1 - b_1 L)(1 - b_2 L)]$$

If both b_1 and b_2 are less than unity in absolute value, we can apply property 5 to obtain

$$y_t = \frac{[a_0/(1 - b_1)] + \sum_{i=0}^{\infty} b_1^i \varepsilon_{t-i}}{1 - b_2 L}$$

Reapply the rule to $a_0/(1 - b_1)$ and to each of the elements in the summation $\Sigma b_1^i \varepsilon_{t-i}$ to obtain the particular solution. If you want to know the actual coefficients of the process, it is preferable to use the method of undetermined coefficients. The beauty of lag operators is that they can be used to denote such particular solutions succinctly. The general model

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

has the particular solution

$$y_t = a_0/A(L) + B(L)\varepsilon_t/A(L)$$

As suggested by (1.82), there is a **forward-looking** solution to any linear difference equation. This text will not make much use of the forward-looking solution since future realizations of stochastic variables are not directly observable. Some of the details of forward-looking solutions can be found at www.cba.ua.edu/~enders.

10. SUMMARY

Time-series econometrics is concerned with the estimation of difference equations containing stochastic components. Originally, time-series models were used for forecasting. Uncovering the dynamic path of a series improves forecasts because the predictable components of the series can be extrapolated into the future. The growing interest in economic dynamics has given a new emphasis to time-series econometrics. Stochastic difference equations arise quite naturally from dynamic economic models. Appropriately estimated equations can be used for the interpretation of economic data and for hypothesis testing.

This introductory chapter focused on methods of "solving" stochastic difference equations. Although iteration can be useful, it is impractical in many circumstances. The solution to a linear difference equation can be divided into two parts: a *particular* solution and a *homogeneous* solution. One complicating factor is that the homogeneous solution is not unique. The *general* solution is a linear combination of the particular solution and all homogeneous solutions. Imposing n initial conditions on the general solution of an nth-order equation yields a unique solution.

The homogeneous portion of a difference equation is a measure of the *disequilibrium* in the initial period(s). The homogeneous equation is especially important in that it yields the characteristic roots; an nth-order equation has n such characteristic roots. If all of the characteristic roots lie within the unit circle, the series will be convergent. As you will see in Chapter 2, there is a direct relationship between the stability conditions and the issue of whether an economic variable is stationary or nonstationary.

The method of undetermined coefficients and the use of lag operators are powerful tools for obtaining the particular solution. The particular solution will be a linear function of the current and past values of the forcing process. In addition, this solution may contain an intercept term and a polynomial function of time. Unit roots and characteristic roots outside of the unit circle require the imposition of an initial condition for the particular solution to be meaningful. Some economic models allow for forward-looking solutions; in such circumstances, anticipated future events have consequences for the present period.

The tools developed in this chapter are aimed at paving the way for the study of time-series econometrics. It is a good idea to work all of the exercises presented below. Characteristic roots, the method of undetermined coefficients, and lag operators will be encountered throughout the remainder of the text.

QUESTIONS AND EXERCISES

1. Consider the difference equation $y_t = a_0 + a_1y_{t-1}$ with the initial condition y_0. Jill solved the difference equation by iterating backward

$$\begin{aligned} y_t &= a_0 + a_1y_{t-1} \\ &= a_0 + a_1(a_0 + a_1y_{t-2}) \\ &= a_0 + a_0a_1 + a_0a_1^2 + \ldots + a_0a_1^{t-1} + a_1^ty_0 \end{aligned}$$

Bill added the homogeneous and particular solutions to obtain $y_t = a_0/(1 - a_1) + a_1^t[y_0 - a_0/(1 - a_1)]$.

   a. Show that the two solutions are identical for $|a_1| < 1$.

   b. Show that for $a_1 = 1$, Jill's solution is equivalent to $y_t = a_0t + y_0$. How would you use Bill's method to arrive at this same conclusion in the case that $a_1 = 1$?

2. The cobweb model in Section 5 assumed *static* price expectations. Consider an alternative formulation called *adaptive expectations*. Let the expected price in t (denoted by p_t^*) be a weighted average of the price in t–1 and the price expectation of the previous period. Formally,

$$p_t^* = \alpha p_{t-1} + (1-\alpha)p_{t-1}^* \qquad 0 < \alpha \leq 1$$

Clearly, when $\alpha = 1$, the static and adaptive expectations schemes are equivalent. An interesting feature of this model is that it can be viewed as a difference equation expressing the expected price as a function of its own lagged value and the forcing variable p_{t-1}.

   a. Find the homogeneous solution for p_t^*.

   b. Use lag operators to find the particular solution. Check your answer by substituting your answer into the original difference equation.

3. Suppose that the money supply process has the form $m_t = m + \rho m_{t-1} + \varepsilon_t$, where m is a constant and $0 < \rho < 1$.

   a. Show that it is possible to express m_{t+n} in terms of the known value m_t and the sequence $\{\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots, \varepsilon_{t+n}\}$.

   b. Suppose that all values of ε_{t+i} for $i > 0$ have a mean value of zero. Explain how you could use your result in part a to forecast the money supply n periods into the future.

4. Find the particular solutions for each of the following:

   a. $y_t = a_1y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$

   b. $y_t = a_1y_{t-1} + \varepsilon_{1t} + \beta\varepsilon_{2t}$ (Hint: The form of the solution is $y_t = \Sigma c_i\varepsilon_{1t-i} + \Sigma d_i\varepsilon_{2t-i}$)

5. The *unit root problem* in time-series econometrics is concerned with characteristic roots that are equal to unity. In order to preview the issue:

   a. Find the homogeneous solution to each of the following (*Hint*: Each has at least one unit root):

      i. $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \varepsilon_t$ **ii.** $y_t = y_{t-2} + \varepsilon_t$

      iii. $y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t$ **iv.** $y_t = y_{t-1} + 0.25y_{t-2} - 0.25y_{t-3} + \varepsilon_t$

   b. Show that each of the backward-looking solutions is not convergent.

   c. Show that Equation i can be written entirely in first differences; that is, $\Delta y_t = 0.5\Delta y_{t-1} + \varepsilon_t$. Find the particular solution for Δy_t. (*Hint*: Find the particular solution for the $\{\Delta y_t\}$ sequence in terms of the $\{\varepsilon_t\}$ sequence.)

   d. Similarly transform the other equations into their first-difference form. Find the backward-looking particular solution, if it exists, for the transformed equations.

   e. Given an initial condition y_0, find the solution for: $y_t = a_0 - y_{t-1} + \varepsilon_t$.

6. A researcher estimated the following relationship for the inflation rate (π_t):

$$\pi_t = -0.05 + 0.7\pi_{t-1} + 0.6\pi_{t-2} + \varepsilon_t$$

   a. Suppose that in periods 0 and 1, the inflation rate was 10 percent and 11 percent, respectively. Find the homogeneous, particular, and general solutions for the inflation rate.

   b. Discuss the shape of the impulse response function. Given that the United States is not headed for runaway inflation, why do you believe that the researcher's equation is poorly estimated?

7. Consider the stochastic process $y_t = a_0 + a_2y_{t-2} + \varepsilon_t$.

   a. Find the homogeneous solution and determine the stability condition.

   b. Find the particular solution using the method of undetermined coefficients.

8. Consider the Cagan (1956) demand for money function in which $m_t - p_t = \alpha - \beta(p_{t+1} - p_t)$.

   a. Show that the backward-looking particular solution for p_t is divergent.

   b. Obtain the forward-looking particular solution for p_t in terms of the $\{m_t\}$ sequence. In forming the general solution, why is it necessary to assume that the money market is in long-run equilibrium?

   c. Find the impact multiplier. How does an increase in m_{t+2} affect p_t? Provide an intuitive explanation of the shape of the entire impulse response function.

9. For each of the following, verify that the posited solution satisfies the difference equation. The symbols c, c_0, and a_0 denote constants.

Equation	Solution
(a) $y_t - y_{t-1} = 0$	$y_t = c$
(b) $y_t - y_{t-1} = a_0$	$y_t = c + a_0t$
(c) $y_t - y_{t-2} = 0$	$y_t = c + c_0(-1)^t$
(d) $y_t - y_{t-2} = \varepsilon_t$	$y_t = c + c_0(-1)^t + \varepsilon_t + \varepsilon_{t-2} + \varepsilon_{t-4} + \ldots$

10. Part 1: For each of the following, determine whether $\{y_t\}$ represents a stable process. Determine whether the characteristic roots are real or imaginary and whether the real parts are positive or negative.

a. $y_t - 1.2y_{t-1} + 0.2y_{t-2}$ **b.** $y_t - 1.2y_{t-1} + 0.4y_{t-2}$

c. $y_t - 1.2y_{t-1} - 1.2y_{t-2}$ **d.** $y_t + 1.2y_{t-1}$

e. $y_t - 0.7y_{t-1} - 0.25y_{t-2} + 0.175y_{t-3} = 0$

[*Hint*: $(x - 0.5)(x + 0.5)(x - 0.7) = x^3 - 0.7x^2 - 0.25x + 0.175$.]

Part 2: Write each of the above equations using lag operators. Determine the characteristic roots of the inverse characteristic equation.

11. Consider the stochastic difference equation

$$y_t = 0.8y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$$

a. Suppose that the initial conditions are such that $y_0 = 0$ and $\varepsilon_0 = \varepsilon_{-1} = 0$. Now suppose that $\varepsilon_1 = 1$. Determine the values y_1 through y_5 by forward iteration.

b. Find the homogeneous and particular solutions.

c. Impose the initial conditions in order to obtain the general solution.

d. Trace out the time path of an ε_t shock on the entire time path of the $\{y_t\}$ sequence.

12. Use equation (1.5) to determine the restrictions on α and β necessary to ensure that the $\{y_t\}$ process is stable.

ENDNOTES

1. Another possibility is to obtain the forward-looking solution; such solutions are discussed in Section 10.
2. Alternatively, you can substitute (1.26) into (1.17). Note that when ε_t is a pure random disturbance, $y_t = a_0 + y_{t-1} + \varepsilon_t$ is called a random walk plus drift model.
3. Any linear equation in the variables x_1 through x_n is homogeneous if it has the form $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$. To obtain the homogeneous portion of (1.10), simply set the intercept term a_0 and the forcing process x_t equal to zero. Hence, the homogeneous equation for (1.10) is $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_n y_{t-n}$.
4. If $b > a$, the demand and supply curves do not intersect in the positive quadrant. The assumption $a > b$ guarantees that the equilibrium price is positive.
5. For example, if the forcing process is $x_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \dots$, the impact multiplier is the partial derivative of y_t with respect to ε_t.

APPENDIX 1.1: Imaginary Roots and de Moivre's Theorem

Consider a second-order difference equation $y_t = a_1 y_{t-1} + a_2 y_{t-2}$ such that the discriminant d is negative [i.e., $d = a_1^2 + 4a_2 < 0$]. From Section 6, we know that the full homogeneous solution can be written in the form

$$y_t^h = A_1\alpha_1^t + A_2\alpha_2^t \qquad \text{(A1.1)}$$

where the two imaginary characteristic roots are

$$\alpha_1 = (a_1 + i\sqrt{-d})/2 \text{ and } \alpha_2 = (a_1 - i\sqrt{-d})/2 \qquad \text{(A1.2)}$$

The purpose of this appendix is to explain how to rewrite and interpret (A1.1) in terms of standard trigonometric functions. You might first want to refresh your memory concerning two useful trig identities. For any two angles θ_1 and θ_2,

$$\sin(\theta_1 + \theta_2) = \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2)$$
$$\cos(\theta_1 + \theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) \qquad \text{(A1.3)}$$

If $\theta_1 = \theta_2$, we can drop subscripts and form

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta)$$
$$\cos(2\theta) = \cos(\theta)\cos(\theta) - \sin(\theta)\sin(\theta) \qquad \text{(A1.4)}$$

The first task is to demonstrate how to express imaginary numbers in the complex plane. Consider Figure A1.1 in which the horizontal axis measures real numbers and the vertical axis measures imaginary numbers. The complex number $a + bi$ can be represented by the point a units from the origin along the horizontal axis and b units from the origin along the vertical axis. It is convenient to represent the distance from the origin by the length of the vector denoted by r. Consider angle θ in triangle $0ab$ and note that $\cos(\theta) = a/r$ and $\sin(\theta) = b/r$. Hence, the lengths a and b can be measured by

$$a = r\cos(\theta) \quad \text{and} \quad b = r\sin(\theta)$$

In terms of (A1.2), we can define $a = a_1/2$ and $b = \sqrt{d}/2$. Thus, the characteristic roots α_1 and α_2 can be written as:

$$\alpha_1 = a + bi = r[\cos(\theta) + i\sin(\theta)]$$
$$\alpha_2 = a - bi = r[\cos(\theta) - i\sin(\theta)] \qquad \text{(A1.5)}$$

The next step is to consider the expressions α_1^t and α_2^t. Begin with the expression α_1^2 and recall that $i^2 = -1$:

$$\alpha_1^2 = \{r[\cos(\theta) + i\sin(\theta)]\}\{r[\cos(\theta) + i\sin(\theta)]\}$$
$$= r^2[\cos(\theta)\cos(\theta) - \sin(\theta)\sin(\theta) + 2i\sin(\theta)\cos(\theta)]$$

From (A1.4),

$$\alpha_1^2 = r^2[\cos(2\theta) + i\sin(2\theta)]$$

If we continue in this fashion, it is straightforward to demonstrate that

$$\alpha_1^t = r^t[\cos(t\theta) + i\sin(t\theta)] \text{ and } \alpha_2^t = r^t[\cos(t\theta) - i\sin(t\theta)]$$

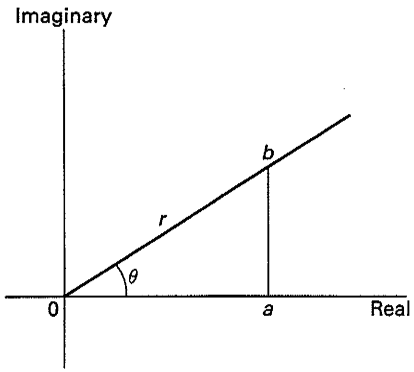


FIGURE A1.1 A Graphical Representation of Complex Numbers

Since y_t^h is a real number and α_1 and α_2 are complex, it follows that A_1 and A_2 must be complex. Although A_1 and A_2 are arbitrary complex numbers, they must have the form

$$A_1 = B_1[\cos(B_2) + i\sin(B_2)] \text{ and } A_2 = B_1[\cos(B_2) - i\sin(B_2)] \quad \text{(A1.7)}$$

where B_1 and B_2 are arbitrary real numbers measured in radians.

In order to calculate $A_1(\alpha_1^t)$, use (A1.6) and (A1.7) to form

$$\begin{aligned} A_1\alpha_1^t &= B_1[\cos(B_2) + i\sin(B_2)]r^t[\cos(t\theta) + i\sin(t\theta)] \\ &= B_1r^t[\cos(B_2)\cos(t\theta) - \sin(B_2)\sin(t\theta) + i\cos(t\theta)\sin(B_2) + i\sin(t\theta)\cos(B_2)] \end{aligned}$$

Using (A1.3), we obtain

$$A_1\alpha_1^t = B_1r^t[\cos(t\theta + B_2) + i\sin(t\theta + B_2)] \quad \text{(A1.8)}$$

You should use the same technique to convince yourself that

$$A_2\alpha_2^t = B_1r^t[\cos(t\theta + B_2) - i\sin(t\theta + B_2)] \quad \text{(A1.9)}$$

Since the homogeneous solution y_t^h is the sum of (A1.8) and (A1.9),

$$\begin{aligned} y_t^h &= B_1r^t[\cos(t\theta + B_2) + i\sin(t\theta + B_2)] + B_1r^t[\cos(t\theta + B_2) - i\sin(t\theta + B_2)] \\ &= 2B_1r^t\cos(t\theta + B_2) \end{aligned} \quad \text{(A1.10)}$$

Since B_1 is arbitrary, the homogeneous solution can be written in terms of the arbitrary constants B_2 and B_3

$$y_t^h = B_3r^t\cos(t\theta + B_2) \quad \text{(A1.11)}$$

Now imagine a circle with a radius of unity superimposed on Figure A1.1. The stability condition is for the distance $r = 0b$ to be less than unity. Hence, in the literature it is said that the stability condition is for the characteristic root(s) to lie within this unit circle.

APPENDIX 1.2: Characteristic Roots in Higher-Order Equations

The characteristic equation to an nth-order difference equation is

$$\alpha^n - a_1\alpha^{n-1} - a_2\alpha^{n-2} \ldots - a_n = 0 \quad \text{(A1.12)}$$

As stated in Section 6, the n values of α which solve this characteristic equation are called the **characteristic roots**. Denote the n solutions by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Given the results in Section 4, the linear combination $A_1\alpha_1^t + A_2\alpha_2^t + \ldots + A_n\alpha_n^t$ is also a solution to (A1.12)

A priori, the characteristic roots can take on any values. There is no restriction that they be real versus complex nor any restriction concerning their sign or magnitude. Consider the possibilities:

1. **All the α_i are real and distinct.** There are several important subcases. First suppose that each value of α_i is less than unity in absolute value. In this case, the homogeneous solution (A1.12) converges since the limit of each α_i^t equals zero as t approaches infinity. For a negative value of α_i,

the expression α_i^t is positive for even values of t and negative for odd values of t. Thus, if any of the α_i are negative (but less than one in absolute value), the solution will tend to exhibit some oscillation. If any of the α_i are greater than unity in absolute value, the solution will diverge.

2. **All of the α_i are real but $m \le n$ of the roots are repeated.** Let the solution be such that $\alpha_1 = \alpha_2 = \ldots = \alpha_m$. Call the single distinct value of this root $\bar{\alpha}$ and let the other n-m roots be denoted by α_{m+1} through α_n. In the case of a second-order equation with a repeated root, you saw that one solution was $A_1\bar{\alpha}^t$ and the other was $A_2t\bar{\alpha}^t$. With m repeated roots, it is easily verified that $t\bar{\alpha}^t, t^2\bar{\alpha}^t, \ldots, t^{m-1}\bar{\alpha}^t$ are also solutions to the homogeneous equation. With m repeated roots, the linear combination of all these solutions is

$$A_1\bar{\alpha}^t + A_2t\bar{\alpha}^t + A_3t^2\bar{\alpha}^t + \ldots + A_mt^{m-1}\bar{\alpha}^t + A_{m+1}\alpha_{m+1}^t + \ldots + A_n\alpha_n^t$$

3. **Some of the roots are complex.** Complex roots (which necessarily come in conjugate pairs) have the form $\alpha_i \pm i\theta$, where α_i and θ are real numbers and i is defined to be $\sqrt{-1}$. For any such pair, a solution to the homogeneous equation is: $A_1(\alpha_1 + i\theta)^t + A_2(\alpha_1 - i\theta)^t$ where A_1 and A_2 are arbitrary constants. Transforming to polar coordinates, the associated two solutions can be written in the form: $\beta_1r^t\cos(\theta t + \beta_2)$ with arbitrary constants β_1 and β_2. Here stability hinges on the magnitude of r^t; if $|r| < 1$, the system converges. However, even if there is convergence, convergence is not direct because the sine and cosine functions impart oscillatory behavior to the time path of y_t. For example, if there are three roots, two of which are complex, the homogeneous solution has the form

$$\beta_1r^t\cos(\theta t + \beta_2) + A_3(\alpha_3)^t$$

Stability of Higher-Order Systems: In practice, it is difficult to find the actual values of the characteristic roots. Unless the characteristic equation is easily factored, it is necessary to use numerical methods to obtain the characteristic roots. Fortunately software packages such as Mathematica, Maple, or Mathcad can easily obtain the characteristic roots of any specific difference equation. At one time it was popular to use the **Schur Theorem** to determine whether all of the roots lie within the unit circle. Rather than calculate all of these determinants, it is often possible to use the simple rules discussed in Section 6. Those of you familiar with matrix algebra may wish to consult the first edition of this text or Samuelson (1941) for the appropriate conditions.