# 20
# TIME-SERIES MODELS

## 20.1 INTRODUCTION

For forecasting purposes, a simple model that *describes* the behavior of a variable (or a set of variables) in terms of past values, without the benefit of a well-developed theory, may well prove quite satisfactory. Researchers have observed that the large simultaneous-equations macroeconomic models constructed in the 1960s frequently have poorer forecasting performance than fairly simple, univariate time-series models based on just a few parameters and compact specifications. It is just this observation that has raised to prominence the univariate time-series forecasting models pioneered by Box and Jenkins (1984).

In this chapter, we introduce some of the tools employed in the analysis of time-series data.[1] Section 20.2 describes stationary stochastic processes. We encountered this body of theory in Chapters 12, 16, and 19, where we discovered that certain assumptions were required to ascribe familiar properties to a time-series of data. We continue that discussion by defining several characteristics of a stationary time-series. The recent literature in macroeconometrics has seen an explosion of studies of nonstationary time series. Nonstationarity mandates a revision of the standard inference tools we have used thus far. In Section 20.3, on nonstationarity and unit roots, we discuss some of these tools. Section 20.4 on cointegration discusses some extensions of regression models that are made necessary when strongly trended, nonstationary variables appear in them.

Some of the concepts to be discussed here were introduced in Section 12.2. Section 12.2 also contains a cursory introduction to the nature of time-series processes. It will be useful to review that material before proceeding with the rest of this chapter. Finally, Sections 15.9.1 on estimation and 15.9.2 and 19.4.3 on stability of dynamic models will be especially useful for the latter sections of this chapter.

[1]Each topic discussed here is the subject of a vast literature with articles and book-length treatments at all levels. For example, two survey papers on the subject of unit roots in economic time-series data, Diebold and Nerlove (1990) and Campbell and Perron (1991) cite between them over 200 basic sources on the subject. The literature on unit roots and cointegration is almost surely the most rapidly moving target in econometrics. Stock's (1994) survey adds hundreds of references to those in the aforementioned surveys and brings the literature up to date as of then. Useful basic references on the subjects of this chapter are Box and Jenkins (1984); Judge et al. (1985); Mills (1990); Granger and Newbold (1996); Granger and Watson (1984); Hendry, Pagan, and Sargan (1984); Geweke (1984); and especially Harvey (1989, 1990); Enders (1995); Hamilton (1994) and Patterson (2000). There are also many survey style and pedagogical articles on these subjects. The aforementioned paper by Diebold and Nerlove is a useful tour guide through some of the literature. We recommend Dickey, Bell, and Miller (1986) and Dickey, Jansen, and Thorton (1991) as well. The latter is an especially clear introduction at a very basic level of the fundamental tools for empirical researchers.

## 20.2 STATIONARY STOCHASTIC PROCESSES

The essential building block for the models to be discussed in this chapter is the **white noise** time-series process,

$$\{\varepsilon_t\}, t = -\infty, +\infty,$$

where each element in the sequence has $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_\varepsilon^2$, and $\text{Cov}[\varepsilon_t, \varepsilon_s] = 0$ for all $s \neq t$. Each element in the series is a random draw from a population with zero mean and constant variance. It is occasionally assumed that the draws are independent or normally distributed, although for most of our analysis, neither assumption will be essential.

A **univariate time-series** model describes the behavior of a variable in terms of its own past values. Consider, for example, the autoregressive disturbance models introduced in Chapter 12,

$$u_t = \rho u_{t-1} + \varepsilon_t. \tag{20-1}$$

Autoregressive disturbances are generally the residual variation in a regression model built up from what may be an elaborate underlying theory, $y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t$. The theory usually stops short of stating what enters the disturbance. But the presumption that some time-series process generates $\mathbf{x}_t$ should extend equally to $u_t$. There are two ways to interpret this simple series. As stated above, $u_t$ equals the previous value of $u_t$ plus an "innovation," $\varepsilon_t$. Alternatively, by manipulating the series, we showed that $u_t$ could be interpreted as an aggregation of the entire history of the $\varepsilon_t$'s.

Occasionally, statistical evidence is convincing that a more intricate process is at work in the disturbance. Perhaps a second-order **autoregression,**

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t, \tag{20-2}$$

better explains the movement of the disturbances in the regression. The model may not arise naturally from an underlying behavioral theory. But in the face of certain kinds of statistical evidence, one might conclude that the more elaborate model would be preferable.[2] This section will describe several alternatives to the AR(1) model that we have relied on in most of the preceding applications.

### 20.2.1 AUTOREGRESSIVE MOVING-AVERAGE PROCESSES

The variable $y_t$ in the model

$$y_t = \mu + \gamma y_{t-1} + \varepsilon_t \tag{20-3}$$

is said to be **autoregressive** (or self-regressive) because under certain assumptions,

$$E[y_t \mid y_{t-1}] = \mu + \gamma y_{t-1}.$$

A more general $p$th-order autoregression or AR($p$) process would be written

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t. \tag{20-4}$$

[2]For example, the estimates of $\varepsilon_t$ computed after a correction for first-order autocorrelation may fail tests of randomness such as the LM (Section 12.7.1) test.

The analogy to the classical regression is clear. Now consider the first order moving average, or MA(1) specification

$$y_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}. \tag{20-5}$$

By writing

$$y_t = \mu + (1 - \theta L)\varepsilon_t$$

or

$$\frac{y_t}{1 - \theta L} = \frac{\mu}{1 - \theta} + \varepsilon_t,^{3}$$

we find that

$$y_t = \frac{\mu}{1 - \theta} - \theta y_{t-1} - \theta^2 y_{t-2} - \cdots + \varepsilon_t.$$

Once again, the effect is to represent $y_t$ as a function of its own past values.

An extremely general model that encompasses (20-4) and (20-5) is the **autoregressive moving average,** or ARMA($p, q$), model:

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}. \tag{20-6}$$

Note the convention that the ARMA($p, q$) process has $p$ autoregressive (lagged dependent-variable) terms and $q$ lagged moving-average terms. Researchers have found that models of this sort with relatively small values of $p$ and $q$ have proved quite effective as forecasting models.

The disturbances $\varepsilon_t$ are labeled the **innovations** in the model. The term is fitting because the only new information that enters the processes in period $t$ is this innovation. Consider, then, the AR(1) process

$$y_t = \mu + \gamma y_{t-1} + \varepsilon_t. \tag{20-7}$$

Either by successive substitution or by using the lag operator, we obtain

$$(1 - \gamma L)y_t = \mu + \varepsilon_t$$

or

$$y_t = \frac{\mu}{1 - \gamma} + \sum_{i=0}^{\infty} \gamma^i \varepsilon_{t-i}.^{4} \tag{20-8}$$

The observed series is a particular type of aggregation of the history of the innovations. The moving average, MA($q$) model,

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} = \mu + D(L)\varepsilon_t, \tag{20-9}$$

is yet another, particularly simple form of aggregation in that only information from the $q$ most recent periods is retained. The general result is that many time-series processes can be viewed either as regressions on lagged values with additive disturbances or as

[3]The lag operator is discussed in Section 19.2.2. Since $\mu$ is a constant, $(1 - \theta L)^{-1}\mu = \mu + \theta\mu + \theta^2\mu + \cdots = \mu/(1 - \theta)$. The lag operator may be set equal to one when it operates on a constant.

[4]See Section 19.3.2 for discussion of models with infinite lag structures.

aggregations of a history of innovations. They differ from one to the next in the form of that aggregation.

More involved processes can be similarly represented in either an autoregressive or moving-average form. (We will turn to the mathematical requirements below.) Consider, for example, the ARMA(2, 1) process,

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \varepsilon_t - \theta\varepsilon_{t-1},$$

which we can write as

$$(1 - \theta L)\varepsilon_t = y_t - \mu - \gamma_1 y_{t-1} - \gamma_2 y_{t-2}.$$

If $|\theta| < 1$, then we can divide both sides of the equation by $(1 - \theta L)$ and obtain

$$\varepsilon_t = \sum_{i=0}^{\infty} \theta^i (y_{t-i} - \mu - \gamma_1 y_{t-i-1} - \gamma_2 y_{t-i-2}).$$

After some tedious manipulation, this equation produces the autoregressive form,

$$y_t = \frac{\mu}{1-\theta} + \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t,$$

where

$$\pi_1 = \gamma_1 - \theta \quad \text{and} \quad \pi_j = -(\theta^j - \gamma_1\theta^{j-1} - \gamma_2\theta^{j-2}), \quad j = 2, 3, \ldots. \tag{20-10}$$

Alternatively, by similar (yet more tedious) manipulation, we would be able to write

$$y_t = \frac{\mu}{1-\gamma_1-\gamma_2} + \left[\frac{1-\theta L}{1-\gamma_1 L - \gamma_2 L^2}\right]\varepsilon_t = \frac{\mu}{1-\gamma_1-\gamma_2} + \sum_{i=0}^{\infty} \delta_i \varepsilon_{t-i}. \tag{20-11}$$

In each case, the weights, $\pi_i$ in the **autoregressive form** and $\delta_i$ in the **moving-average form** are complicated functions of the original parameters. But nonetheless, each is just an alternative representation of the same time-series process that produces the current value of $y_t$. This result is a fundamental property of certain time series. We will return to the issue after we formally define the assumption that we have used at several steps above that allows these transformations.

### 20.2.2 STATIONARITY AND INVERTIBILITY

At several points in the preceding, we have alluded to the notion of **stationarity,** either directly or indirectly by making certain assumptions about the parameters in the model. In Section 12.3.2, we characterized an AR(1) disturbance process

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

as stationary if $|\rho| < 1$ and $\varepsilon_t$ is **white noise.** Then

$$\begin{aligned} E[u_t] &= 0 \quad \text{for all } t, \\ \text{Var}[u_t] &= \frac{\sigma_\varepsilon^2}{1-\rho^2}, \\ \text{Cov}[u_t, u_s] &= \frac{\rho^{|t-s|}\sigma_\varepsilon^2}{1-\rho^2}. \end{aligned} \tag{20-12}$$

If $|\rho| \geq 1$, then the variance and covariances are undefined.

In the following, we use $\varepsilon_t$ to denote the white noise innovations in the process. The ARMA$(p, q)$ process will be denoted as in (20-6).

**DEFINITION 20.1 Covariance Stationarity**

*A stochastic process $y_t$ is **weakly stationary** or **covariance stationary** if it satisfies the following requirements:*[5]

1. $E[y_t]$ *is independent of* $t$.
2. $\text{Var}[y_t]$ *is a finite, positive constant, independent of* $t$.
3. $\text{Cov}[y_t, y_s]$ *is a finite function of* $|t - s|$, *but not of* $t$ *or* $s$.

The third requirement is that the covariance between observations in the series is a function only of how far apart they are in time, not the time at which they occur. These properties clearly hold for the AR(1) process immediately above. Whether they apply for the other models we have examined remains to be seen.

We define the **autocovariance at lag *k*** as

$$\lambda_k = \text{Cov}[y_t, y_{t-k}].$$

Note that

$$\lambda_{-k} = \text{Cov}[y_t, y_{t+k}] = \lambda_k.$$

Stationarity implies that autocovariances are a function of $k$, but not of $t$. For example, in (20-12), we see that the autocovariances of the AR(1) process $y_t = \mu + \gamma y_{t-1} + \varepsilon_t$ are

$$\text{Cov}[y_t, y_{t-k}] = \frac{\gamma^k \sigma_\varepsilon^2}{1 - \gamma^2}, \quad k = 0, 1 \ldots. \tag{20-13}$$

If $|\gamma| < 1$, then this process is stationary. For any MA$(q)$ series,

$$\begin{aligned} y_t &= \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \\ E[y_t] &= \mu + E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] - \cdots - \theta_q E[\varepsilon_{t-q}] = \mu, \\ \text{Var}[y_t] &= \left(1 + \theta_1^2 + \cdots + \theta_q^2\right) \sigma_\varepsilon^2, \\ \text{Cov}[y_t, y_{t-1}] &= \left(-\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \cdots + \theta_{q-1}\theta_q\right) \sigma_\varepsilon^2, \end{aligned} \tag{20-14}$$

and so on until

$$\begin{aligned} \text{Cov}[y_t, y_{t-(q-1)}] &= \left[-\theta_{q-1} + \theta_1\theta_q\right] \sigma_\varepsilon^2, \\ \text{Cov}[y_t, y_{t-q}] &= -\theta_q \sigma_\varepsilon^2, \end{aligned}$$

[5] *Strong* stationarity requires that the joint distribution of all sets of observations $(y_t, y_{t-1}, \ldots)$ be invariant to when the observations are made. For practical purposes in econometrics, this statement is a theoretical fine point. Although weak stationary suffices for our applications, we would not normally analyze weakly stationary time series that were not strongly stationary as well. Indeed, we often go even beyond this step and assume joint normality.

and, for lags greater than $q$, the autocovariances are zero. It follows, therefore, that finite moving-average processes are stationary regardless of the values of the parameters. The MA(1) process $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$ is an important special case that has $\text{Var}[y_t] = (1+\theta^2)\sigma_\varepsilon^2$, $\lambda_1 = -\theta\sigma_\varepsilon^2$, and $\lambda_k = 0$ for $|k| > 1$.

For the AR(1) process, the stationarity requirement is that $|\gamma| < 1$, which in turn, implies that the variance of the moving average representation in (20-8) is finite. Consider the AR(2) process

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \varepsilon_t.$$

Write this equation as

$$C(L)y_t = \mu + \varepsilon_t,$$

where

$$C(L) = 1 - \gamma_1 L - \gamma_2 L^2.$$

Then, if it is possible, we invert this result to produce

$$y_t = [C(L)]^{-1}(\mu + \varepsilon_t).$$

Whether the inversion of the polynomial in the lag operator leads to a convergent series depends on the values of $\gamma_1$ and $\gamma_2$. If so, then the moving-average representation will be

$$y_t = \sum_{i=0}^{\infty} \delta_i(\mu + \varepsilon_{t-i})$$

so that

$$\text{Var}[y_t] = \sum_{i=0}^{\infty} \delta_i^2 \sigma_\varepsilon^2.$$

Whether this result is finite or not depends on whether the series of $\delta_i$s is exploding or converging. For the AR(2) case, the series converges if $|\gamma_2| < 1$, $\gamma_1 + \gamma_2 < 1$, and $\gamma_2 - \gamma_1 < 1$.[6]

For the more general case, the autoregressive process is stationary if the roots of the **characteristic equation,**

$$C(z) = 1 - \gamma_1 z - \gamma_2 z^2 - \cdots - \gamma_p z^p = 0,$$

have modulus greater than one, or "lie outside the unit circle."[7] It follows that if a stochastic process is stationary, it has an infinite moving-average representation (and, if not, it does not). The AR(1) process is the simplest case. The characteristic equation is

$$C(z) = 1 - \gamma z = 0,$$

[6]This requirement restricts $(\gamma_1, \gamma_2)$ to within a triangle with points at $(2, -1)$, $(-2, -1)$, and $(0, 1)$.

[7]The roots may be complex. (See Sections 15.9.2 and 19.4.3.) They are of the form $a \pm bi$, where $i = \sqrt{-1}$. The unit circle refers to the two-dimensional set of values of $a$ and $b$ defined by $a^2 + b^2 = 1$, which defines a circle centered at the origin with radius 1.

and its single root is $1/\gamma$. This root lies outside the unit circle if $|\gamma| < 1$, which we saw earlier.

Finally, consider the inversion of the moving-average process in (20-9) and (20-10). Whether this inversion is possible depends on the coefficients in $D(L)$ in the same fashion that stationarity hinges on the coefficients in $C(L)$. This counterpart to stationarity of an autoregressive process is called **invertibility.** For it to be possible to invert a moving-average process to produce an autoregressive representation, the roots of $D(L) = 0$ must be outside the unit circle. Notice, for example, that in (20-5), the inversion of the moving-average process is possible only if $|\theta| < 1$. Since the characteristic equation for the MA(1) process is $1 - \theta L = 0$, the root is $1/\theta$, which must be larger than one.

If the roots of the characteristic equation of a moving-average process all lie outside the unit circle, then the series is said to be invertible. Note that invertibility has no bearing on the stationarity of a process. All moving-average processes with finite coefficients are stationary. Whether an ARMA process is stationary or not depends only on the AR part of the model.

### 20.2.3 AUTOCORRELATIONS OF A STATIONARY STOCHASTIC PROCESS

The function

$$\lambda_k = \text{Cov}[y_t, y_{t-k}]$$

is called the **autocovariance function** of the process $y_t$. The **autocorrelation function,** or **ACF,** is obtained by dividing by the variance $\lambda_0$ to obtain

$$\rho_k = \frac{\lambda_k}{\lambda_0}, \quad -1 \leq \rho_k \leq 1.$$

For a stationary process, the ACF will be a function of $k$ and the parameters of the process. The ACF is a useful device for describing a time-series process in much the same way that the moments are used to describe the distribution of a random variable. One of the characteristics of a stationary stochastic process is an autocorrelation function that either abruptly drops to zero at some finite lag or eventually tapers off to zero. The AR(1) process provides the simplest example, since

$$\rho_k = \gamma^k,$$

which is a geometric series that either declines monotonically from $\rho_0 = 1$ if $\gamma$ is positive or with a damped sawtooth pattern if $\gamma$ is negative. Note as well that for the process $y_t = \gamma y_{t-1} + \varepsilon_t$,

$$\rho_k = \gamma \rho_{k-1}, \quad k \geq 1,$$

which bears a noteworthy resemblance to the process itself.

For higher-order autoregressive series, the autocorrelations may decline monotonically or may progress in the fashion of a damped sine wave.[8] Consider, for example, the second-order autoregression, where we assume without loss of generality that $\mu = 0$

[8]The behavior is a function of the roots of the characteristic equation. This aspect is discussed in Section 15.9 and especially 15.9.3.

(since we are examining second moments in deviations from the mean):

$$y_t = \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \varepsilon_t.$$

If the process is stationary, then $\text{Var}[y_t] = \text{Var}[y_{t-s}]$ for all $s$. Also, $\text{Var}[y_t] = \text{Cov}[y_t, y_t]$, and $\text{Cov}[\varepsilon_t, y_{t-s}] = 0$ if $s > 0$. These relationships imply that

$$\lambda_0 = \gamma_1 \lambda_1 + \gamma_2 \lambda_2 + \sigma_\varepsilon^2.$$

Now, using additional lags, we find that

$$\lambda_1 = \gamma_1 \lambda_0 + \gamma_2 \lambda_1$$

and

$$\lambda_2 = \gamma_1 \lambda_1 + \gamma_2 \lambda_0. \qquad \textbf{(20-15)}$$

These three equations provide the solution:

$$\lambda_0 = \sigma_\varepsilon^2 \frac{[(1-\gamma_2)/(1+\gamma_2)]}{(1-\gamma_1^2-\gamma_2^2)}.$$

The variance is unchanging, so we can divide throughout by $\lambda_0$ to obtain the relationships for the autocorrelations,

$$\rho_1 = \gamma_1 \rho_0 + \gamma_2 \rho_1.$$

Since $\rho_0 = 1$, $\rho_1 = \gamma_1/(1-\gamma_2)$. Using the same procedure for additional lags, we find that

$$\rho_2 = \gamma_1 \rho_1 + \gamma_2,$$

so $\rho_2 = \gamma_1^2/(1-\gamma_2) + \gamma_2$. Generally, then, for lags of two or more,

$$\rho_k = \gamma_1 \rho_{k-1} + \gamma_2 \rho_{k-2}.$$

Once again, the autocorrelations follow the same difference equation as the series itself. The behavior of this function depends on $\gamma_1$, $\gamma_2$, and $k$, although not in an obvious way. The inherent behavior of the autocorrelation function can be deduced from the characteristic equation.[9] For the second-order process we are examining, the autocorrelations are of the form

$$\rho_k = \phi_1(1/z_1)^k + \phi_2(1/z_2)^k,$$

where the two roots are[10]

$$1/z = \tfrac{1}{2}\left[\gamma_1 \pm \sqrt{\gamma_1^2 + 4\gamma_2}\right].$$

If the two roots are real, then we know that their reciprocals will be less than one in absolute value, so that $\rho_k$ will be the sum of two terms that are decaying to zero. If the two roots are complex, then $\rho_k$ will be the sum of two terms that are oscillating in the form of a damped sine wave.

[9]The set of results that we would use to derive this result are exactly those we used in Section 19.4.3 to analyze the stability of a dynamic equation, which makes sense, of course, since the equation linking the autocorrelations is a simple difference equation.

[10]We used the device in Section 19.4.4 to find the characteristic roots. For a second-order equation, the quadratic is easy to manipulate.

Applications that involve autoregressions of order greater than two are relatively unusual. Nonetheless, higher-order models can be handled in the same fashion. For the AR($p$) process

$$y_t = \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t,$$

the autocovariances will obey the **Yule–Walker equations**

$$\lambda_0 = \gamma_1 \lambda_1 + \gamma_2 \lambda_2 + \cdots + \gamma_p \lambda_p + \sigma_\varepsilon^2,$$
$$\lambda_1 = \gamma_1 \lambda_0 + \gamma_2 \lambda_1 + \cdots + \gamma_p \lambda_{p-1},$$

and so on. The autocorrelations will once again follow the same difference equation as the original series,

$$\rho_k = \gamma_1 \rho_{k-1} + \gamma_2 \rho_{k-2} + \cdots + \gamma_p \rho_{k-p}.$$

The ACF for a moving-average process is very simple to obtain. For the first-order process,

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1},$$
$$\lambda_0 = (1 + \theta^2)\sigma_\varepsilon^2,$$
$$\lambda_1 = -\theta \sigma_\varepsilon^2,$$

then $\lambda_k = 0$ for $k > 1$. Higher-order processes appear similarly. For the MA(2) process, by multiplying out the terms and taking expectations, we find that

$$\lambda_0 = \left(1 + \theta_1^2 + \theta_2^2\right)\sigma_\varepsilon^2,$$
$$\lambda_1 = (-\theta_1 + \theta_1\theta_2)\sigma_\varepsilon^2,$$
$$\lambda_2 = -\theta_1 \sigma_\varepsilon^2,$$
$$\lambda_k = 0, \quad k > 2.$$

The pattern for the general MA($q$) process $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$ is analogous. The signature of a moving-average process is an autocorrelation function that abruptly drops to zero at one lag past the order of the process. As we will explore below, this sharp distinction provides a statistical tool that will help us distinguish between these two types of processes empirically.

The mixed process, ARMA($p$, $q$), is more complicated since it is a mixture of the two forms. For the ARMA(1, 1) process

$$y_t = \gamma y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1},$$

the Yule–Walker equations are

$$\lambda_0 = E[y_t(\gamma y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})] = \gamma \lambda_1 + \sigma_\varepsilon^2 - \sigma_\varepsilon^2(\theta\gamma - \theta^2),$$
$$\lambda_1 = \gamma \lambda_0 - \theta \sigma_\varepsilon^2,$$

and

$$\lambda_k = \gamma \lambda_{k-1}, \quad k > 1.$$

The general characteristic of ARMA processes is that when the moving-average component is of order $q$, then in the series of autocorrelations there will be an initial $q$ terms that are complicated functions of both the AR and MA parameters, but after $q$ periods,

$$\rho_k = \gamma_1 \rho_{k-1} + \gamma_2 \rho_{k-2} + \cdots + \gamma_p \rho_{k-p}, \quad k > q.$$

### 20.2.4 PARTIAL AUTOCORRELATIONS OF A STATIONARY STOCHASTIC PROCESS

The autocorrelation function ACF($k$) gives the gross correlation between $y_t$ and $y_{t-k}$. But as we saw in our analysis of the classical regression model in Section 3.4, a gross correlation such as this one can mask a completely different underlying relationship. In this setting, we observe, for example, that a correlation between $y_t$ and $y_{t-2}$ could arise primarily because both variables are correlated with $y_{t-1}$. Consider the AR(1) process $y_t = \gamma y_{t-1} + \varepsilon_t$. The second gross autocorrelation is $\rho_2 = \gamma^2$. But in the same spirit, we might ask what is the correlation between $y_t$ and $y_{t-2}$ *net of the intervening effect of* $y_{t-1}$? In this model, if we remove the effect of $y_{t-1}$ from $y_t$, then only $\varepsilon_t$ remains, and this disturbance is uncorrelated with $y_{t-2}$. We would conclude that the **partial autocorrelation** between $y_t$ and $y_{t-2}$ in this model is zero.

**DEFINITION 20.2 Partial Autocorrelation Coefficient**

*The partial correlation between $y_t$ and $y_{t-k}$ is the simple correlation between $y_{t-k}$ and $y_t$ minus that part explained linearly by the intervening lags. That is,*

$$\rho_k^* = \text{Corr}[y_t - E^*(y_t \mid y_{t-1}, \ldots, y_{t-k+1}), y_{t-k}],$$

*where $E^*(y_t \mid y_{t-1}, \ldots, y_{t-k+1})$ is the minimum mean-squared error predictor of $y_t$ by $y_{t-1}, \ldots, y_{t-k+1}$.*

The function $E^*(.)$ might be the linear regression if the conditional mean happened to be linear, but it might not. The optimal *linear* predictor *is* the linear regression, however, so what we have is

$$\rho_k^* = \text{Corr}[y_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \cdots - \beta_{k-1} y_{t-k+1}, y_{t-k}],$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_{k-1}] = \{\text{Var}[y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}]\}^{-1} \times \text{Cov}[y_t, (y_{t-1}, y_{t-2}, \ldots, y_{t-k+1})]'$. This equation will be recognized as a vector of regression coefficients. As such, what we are computing here (of course) is the correlation between a vector of residuals and $y_{t-k}$. There are various ways to formalize this computation [see, e.g., Enders (1995, pp. 82–85)]. One intuitively appealing approach is suggested by the equivalent definition (which is also a prescription for computing it), as follows.

**DEFINITION 20.3 Partial Autocorrelation Coefficient**

*The partial correlation between $y_t$ and $y_{t-k}$ is the last coefficient in the linear projection of $y_t$ on $[y_{t-1}, y_{t-2}, \ldots, y_{t-k}]$,*

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \\ \rho_k^* \end{bmatrix} = \begin{bmatrix} \lambda_0 & \lambda_1 & \cdots & \lambda_{k-2} & \lambda_{k-1} \\ \lambda_1 & \lambda_0 & \cdots & \lambda_{k-3} & \lambda_{k-2} \\ & \cdots & \vdots & \cdots & \\ \lambda_{k-1} & \lambda_{k-2} & \cdots & \lambda_1 & \lambda_0 \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix}.$$

As before, there are some distinctive patterns for particular time-series processes. Consider first the autoregressive processes,

$$y_t = \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t.$$

We are interested in the last coefficient in the projection of $y_t$ on $y_{t-1}$, then on $[y_{t-1}, y_{t-2}]$, and so on. The first of these is the simple regression coefficient of $y_t$ on $y_{t-1}$, so

$$\rho_1^* = \frac{\text{Cov}[y_t, y_{t-1}]}{\text{Var}[y_{t-1}]} = \frac{\lambda_1}{\lambda_0} = \rho_1.$$

The first partial autocorrelation coefficient for any process equals the first autocorrelation coefficient.

Without doing the messy algebra, we also observe that for the AR($p$) process, $\rho_1^*$ is a mixture of all the $\gamma$ coefficients. Of course, if $p$ equals 1, then $\rho_1^* = \rho_1 = \gamma$. For the higher-order processes, the autocorrelations are likewise mixtures of the autoregressive coefficients until we reach $\rho_p^*$. In view of the form of the AR($p$) model, the last coefficient in the linear projection on $p$ lagged values is $\gamma_p$. Also, we can see the signature pattern of the AR($p$) process, any additional partial autocorrelations must be zero, because they will be simply $\rho_k^* = \text{Corr}[\varepsilon_t, y_{t-k}] = 0$ if $k > p$.

Combining results thus far, we have the characteristic pattern for an autoregressive process. The ACF, $\rho_k$, will gradually decay to zero, either monotonically if the characteristic roots are real or in a sinusoidal pattern if they are complex. The PACF, $\rho_k^*$, will be irregular out to lag $p$, when they abruptly drop to zero and remain there.

The moving-average process has the mirror image of this pattern. We have already examined the ACF for the MA($q$) process; it has $q$ irregular spikes, then it falls to zero and stays there. For the PACF, write the model as

$$y_t = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)\varepsilon_t.$$

If the series is invertible, which we will assume throughout, then we have

$$\frac{y_t}{1 - \theta_1 L - \cdots - \theta_q L^q} = \varepsilon_t,$$

or

$$\begin{aligned} y_t &= \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \varepsilon_t \\ &= \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t. \end{aligned}$$

The autoregressive form of the MA($q$) process has an infinite number of terms, which means that the PACF will not fall off to zero the way that the PACF of the AR process does. Rather, the PACF of an MA process will resemble the ACF of an AR process. For example, for the MA(1) process $y_t = \varepsilon_t - \theta\varepsilon_{t-1}$, the AR representation is

$$y_t = \theta y_{t-1} + \theta^2 y_{t-2} + \cdots + \varepsilon_t,$$

which is the familiar form of an AR(1) process. Thus, the PACF of an MA(1) process is identical to the ACF of an AR(1) process, $\rho_k^* = \theta^k$.

The ARMA($p, q$) is a mixture of the two types of processes, so its ACF and PACF are likewise mixtures of the two forms discussed above. Generalities are difficult to

draw, but normally, the ACF of an ARMA process will have a few distinctive spikes in the early lags corresponding to the number of MA terms, followed by the characteristic smooth pattern of the AR part of the model. High-order MA processes are relatively uncommon in general, and high-order AR processes (greater than two) seem primarily to arise in the form of the nonstationary processes described in the next section. For a stationary process, the workhorses of the applied literature are the (2, 0) and (1, 1) processes. For the ARMA(1, 1) process, both the ACF and the PACF will display a distinctive spike at lag 1 followed by an exponentially decaying pattern thereafter.

### 20.2.5 MODELING UNIVARIATE TIME SERIES

The preceding discussion is largely descriptive. There is no underlying economic theory that states *why* a compact ARMA($p$, $q$) representation should adequately describe the movement of a given economic time series. Nonetheless, as a methodology for building forecasting models, this set of tools and its empirical counterpart have proved as good as and even superior to much more elaborate specifications (perhaps to the consternation of the builders of large macroeconomic models).[11] Box and Jenkins (1984) pioneered a forecasting framework based on the preceding that has been used in a great many fields and that has, certainly in terms of numbers of applications, largely supplanted the use of large integrated econometric models.

Box and Jenkins's approach to modeling a stochastic process can be motivated by the following.

**THEOREM 20.1 Wold's Decomposition Theorem**

*Every zero mean covariance stationary stochastic process can be represented in the form*

$$y_t = E^*[y_t \mid y_{t-1}, y_{t-2}, \ldots, t_{t-p}] + \sum_{i=0}^{\infty} \pi_i \varepsilon_{t-i},$$

*where $\varepsilon_t$ is white noise, $\pi_0 = 1$, and the weights are* ***square summable****—that is,*

$$\sum_{i=1}^{\infty} \pi_i^2 < \infty$$

*— $E^*[y_t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-p}]$ is the optimal linear predictor of $y_t$ based on its lagged values, and the predictor $E_t^*$ is uncorrelated with $\varepsilon_{t-i}$.*

Thus, the theorem decomposes the process generating $y_t$ into

$$E_t^* = E^*[y_t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-p}] = \text{the } \textbf{linearly deterministic component}$$

[11] This observation can be overstated. Even the most committed advocate of the Box–Jenkins methods would concede that an ARMA model of, for example, housing starts will do little to reveal the link between the interest rate policies of the Federal Reserve and their variable of interest. That is, the *covariation* of economic variables remains as interesting as ever.

and

$$\sum_{i=0}^{\infty} \pi_i \varepsilon_{t-i} = \text{the } \textbf{linearly indeterministic component.}$$

The theorem states that for any stationary stochastic process, for a given choice of $p$, there is a Wold representation of the stationary series

$$y_t = \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=0}^{\infty} \pi_i \varepsilon_{t-i}.$$

Note that for a specific ARMA($P$, $Q$) process, if $p \geq P$, then $\pi_i = 0$ for $i > Q$. For practical purposes, the problem with the Wold representation is that we cannot estimate the infinite number of parameters needed to produce the full right-hand side, and, of course, $P$ and $Q$ are unknown. The compromise, then, is to base an estimate of the representation on a model with a finite number of moving-average terms. We can seek the one that best fits the data in hand.

It is important to note that neither the ARMA representation of a process nor the Wold representation is unique. In general terms, suppose that the process generating $y_t$ is

$$\Gamma(L) y_t = \Theta(L) \varepsilon_t.$$

We assume that $\Gamma(L)$ is finite but $\Theta(L)$ need not be. Let $\Phi(L)$ be some other polynomial in the lag operator with roots that are outside the unit circle. Then

$$\left[\frac{\Phi(L)}{\Gamma(L)}\right] \Gamma(L) y_t = \left[\frac{\Phi(L)}{\Gamma(L)}\right] \Theta(L) \varepsilon_t$$

or

$$\Phi(L) y_t = \Pi(L) \varepsilon_t.$$

The new representation is fully equivalent to the old one, but it might have a different number of autoregressive parameters, which is exactly the point of the Wold decomposition. The implication is that part of the model-building process will be to determine the lag structures. Further discussion on the methodology is given by Box and Jenkins (1984).

The Box–Jenkins approach to modeling stochastic processes consists of the following steps:

1. Satisfactorily transform the data so as to obtain a stationary series. This step will usually mean taking first differences, logs, or both to obtain a series whose autocorrelation function eventually displays the characteristic exponential decay of a stationary series.
2. Estimate the parameters of the resulting ARMA model, generally by nonlinear least squares.
3. Generate the set of residuals from the estimated model and verify that they satisfactorily resemble a white noise series. If not, respecify the model and return to step 2.
4. The model can now be used for forecasting purposes.

Space limitations prevent us from giving a full presentation of the set of techniques. Since this methodology has spawned a mini-industry of its own, however, there is no shortage of book length analyses and prescriptions to which the reader may refer. Five to consider are the canonical source, Box and Jenkins (1984), Granger and Newbold (1986), Mills (1993), Enders (1995) and Patterson (2000). Some of the aspects of the estimation and analysis steps do have broader relevance for our work here, so we will continue to examine them in some detail.

### 20.2.6 ESTIMATION OF THE PARAMETERS OF A UNIVARIATE TIME SERIES

The broad problem of regression estimation with time series data, which carries through to all the discussions of this chapter, is that the consistency and asymptotic normality results that we derived based on random sampling will no longer apply. For example, for a stationary series, we have assumed that $\text{Var}[y_t] = \lambda_0$ regardless of $t$. But we have yet to establish that an estimated variance,

$$c_0 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2,$$

will converge to $\lambda_0$, or anything else for that matter. It is necessary to assume that the process is **ergodic.** (We first encountered this assumption in Section 12.4.1—see Definition 12.3.) Ergodicity is a crucial element of our theory of estimation. When a time series has this property (with stationarity), then we can consider estimation of parameters in a meaningful sense. If the process is stationary and ergodic then, by the Ergodic Theorem (Theorems 12.1 and 12.2) moments such as $\bar{y}$ and $c_0$ converge to their population counterparts $\mu$ and $\lambda_0$.[12] The essential component of the condition is one that we have met at many points in this discussion, that autocovariances must decline sufficiently rapidly as the separation in time increases. It is possible to construct theoretical examples of processes that are stationary but not ergodic, but for practical purposes, a stationarity assumption will be sufficient for us to proceed with estimation. For example, in our models of stationary processes, if we assume that $\varepsilon_t \sim N[0, \sigma^2]$, which is common, then the stationary processes are ergodic as well.

Estimation of the parameters of a time-series process must begin with a determination of the type of process that we have in hand. (Box and Jenkins label this the **identification** step. But identification is a term of art in econometrics, so we will steer around that admittedly standard name.) For this purpose, the empirical estimates of the autocorrelation and partial autocorrelation functions are useful tools.

The sample counterpart to the ACF is the **correlogram,**

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

A plot of $r_k$ against $k$ provides a description of a process and can be used to help discern what type of process is generating the data. The sample PACF is the counterpart to the

[12]The formal conditions for ergodicity are quite involved; see Davidson and MacKinnon (1993) or Hamilton (1994, Chapter 7).

ACF, but net of the intervening lags; that is,

$$r_k^* = \frac{\sum_{t=k+1}^{T} y_t^* y_{t-k}^*}{\sum_{t=k+1}^{T} (y_{t-k}^*)^2},$$

where $y_t^*$ and $y_{t-k}^*$ are residuals from the regressions of $y_t$ and $y_{t-k}$ on $[1, y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}]$. We have seen this at many points before; $r_k^*$ is simply the last linear least squares regression coefficient in the regression of $y_t$ on $[1, y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}, y_{t-k}]$. Plots of the ACF and PACF of a series are usually presented together. Since the sample estimates of the autocorrelations and partial autocorrelations are not likely to be identically zero even when the population values are, we use diagnostic tests to discern whether a time series appears to be nonautocorrelated.[13] Individual sample autocorrelations will be approximately distributed with mean zero and variance $1/T$ under the hypothesis that the series is white noise. The Box–Pierce (1970) statistic

$$Q = T\sum_{k=1}^{p} r_k^2$$

is commonly used to test whether a series is white noise. Under the null hypothesis that the series is white noise, $Q$ has a limiting chi-squared distribution with $p$ degrees of freedom. A refinement that appears to have better finite-sample properties is the Ljung–Box (1979) statistic,

$$Q' = T(T+2)\sum_{k=1}^{p} \frac{r_k^2}{T-k}.$$

The limiting distribution of $Q'$ is the same as that of $Q$.

The process of finding the appropriate specification is essentially trial and error. An initial specification based on the sample ACF and PACF can be found. The parameters of the model can then be estimated by least squares. For pure AR($p$) processes, the estimation step is simple. The parameters can be estimated by linear least squares. If there are moving-average terms, then linear least squares is inconsistent, but the parameters of the model can be fit by nonlinear least squares. Once the model has been estimated, a set of residuals is computed to assess the adequacy of the specification. In an AR model, the residuals are just the deviations from the regression line.

The adequacy of the specification can be examined by applying the foregoing techniques to the estimated residuals. If they appear satisfactorily to mimic a white noise process, then analysis can proceed to the forecasting step. If not, a new specification should be considered.

***Example 20.1 ACF and PACF for a Series of Bond Yields***

Appendix Table F20.1 lists 5 years of monthly averages of the yield on a Moody's Aaa rated corporate bond. The series is plotted in Figure 20.1. From the figure, it would appear that stationarity may not be a reasonable assumption. We will return to this question below. The ACF and PACF for the original series are shown in Table 20.1, with the diagnostic statistics discussed earlier.

The plots appear to be consistent with an AR(2) process, although the ACF at longer lags seems a bit more persistent than might have been expected. Once again, this condition

[13] The LM test discussed in Section 12.7.1 is one of these.

may indicate that the series is not stationary. Maintaining that assumption for the present, we computed the residuals from the AR(2) model and subjected them to the same tests as the original series. The coefficients of the AR(2) model are 1.1566 and −0.2083, which also satisfy the restrictions for stationarity given in Section 20.2.2. Despite the earlier suggestions, the residuals do appear to resemble a white noise series (Table 20.2).

**FIGURE 20.1** Monthly Data on Bond Yields.

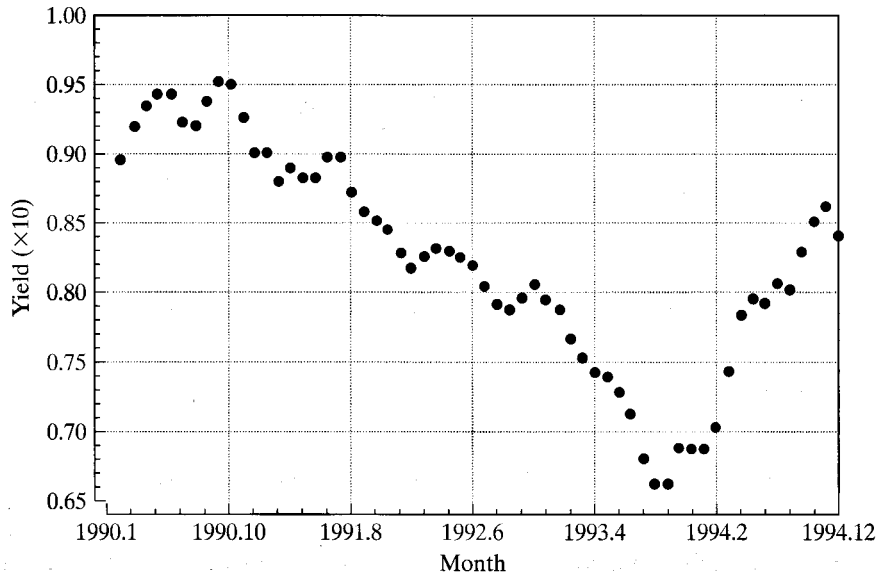



**TABLE 20.1** ACF and PACF for Bond Yields

Time-series identification for YIELD
Box–Pierce statistic = 323.0587
Degrees of freedom = 14
Significance level = 0.0000
Box–Ljung Statistic = 317.4389
Degrees of freedom = 14
Significance level = 0.0000
◆ → |coefficient| > 2/sqrt($N$) or > 95% significant

| | *Autocorrelation Function* | | | | *Partial Autocorrelations* | | | |
|---|---|---|---|---|---|---|---|---|
| *Lag* | | *−1* | *0* | *+1* | *Box–Pierce* | | *−1* | *0* | *+1* |
| 1 | 0.970◆ | | | | 56.42◆ | 0.970◆ | | | |
| 2 | 0.908◆ | | | | 105.93◆ | −0.573◆ | | | |
| 3 | 0.840◆ | | | | 148.29◆ | 0.157 | | | |
| 4 | 0.775◆ | | | | 184.29◆ | −0.043 | | | |
| 5 | 0.708◆ | | | | 214.35◆ | −0.309◆ | | | |
| 6 | 0.636◆ | | | | 238.65◆ | −0.024 | | | |
| 7 | 0.567◆ | | | | 257.93◆ | −0.037 | | | |
| 8 | 0.501◆ | | | | 272.97◆ | 0.059 | | | |
| 9 | 0.439◆ | | | | 284.51◆ | −0.068 | | | |
| 10 | 0.395◆ | | | | 293.85◆ | 0.216 | | | |
| 11 | 0.370◆ | | | | 302.08◆ | −0.180 | | | |
| 12 | 0.354◆ | | | | 309.58◆ | 0.048 | | | |
| 13 | 0.339◆ | | | | 316.48◆ | 0.162 | | | |
| 14 | 0.331◆ | | | | 323.06◆ | 0.171 | | | |

**TABLE 20.2 ACF and PACF for Residuals**

Time-series identification for U
Box–Pierce statistic = 13.7712
Significance level = 0.4669
Box–Ljung statistic = 16.1336
Significance level = 0.3053
◆ → |coefficient| > 2/sqrt($N$) or > 95% significant

| | *Autocorrelation Function* | | | | *Partial Autocorrelations* | | |
|---|---|---|---|---|---|---|---|
| *Lag* | | *−1* | *0* | *Box–Pierce* | | *−1* | *0* |
| 1 | 0.154 | | ■ | 1.38 | 0.154 | | ■ |
| 2 | −0.147 | ■ | | 2.64 | −0.170 | ■ | |
| 3 | −0.207 | ■ | | 5.13 | −0.179 | ■ | |
| 4 | 0.161 | | ■ | 6.64 | 0.183 | | ■ |
| 5 | 0.117 | | ■ | 7.43 | 0.068 | | ▮ |
| 6 | 0.114 | | ■ | 8.18 | 0.094 | | ▮ |
| 7 | −0.110 | ▮ | | 8.89 | −0.066 | ▮ | |
| 8 | 0.041 | | ▮ | 8.99 | 0.125 | | ■ |
| 9 | −0.168 | ■ | | 10.63 | −0.258 | ■ | |
| 10 | 0.014 | | ▮ | 10.64 | 0.035 | | ▮ |
| 11 | −0.016 | ▮ | | 10.66 | 0.015 | | ▮ |
| 12 | −0.009 | ▮ | | 10.66 | −0.089 | ▮ | |
| 13 | −0.195 | ■ | | 12.87 | −0.166 | ■ | |
| 14 | −0.125 | ■ | | 13.77 | 0.132 | | ■ |

### 20.2.7 THE FREQUENCY DOMAIN

For the analysis of macroeconomic flow data such as output and consumption, and aggregate economic index series such as the price level and the rate of unemployment, the tools described in the previous sections have proved quite satisfactory. The low frequency of observation (yearly, quarterly, or, occasionally, monthly) and very significant aggregation (both across time and of individuals) make these data relatively smooth and straightforward to analyze. Much contemporary economic analysis, especially financial econometrics, has dealt with more disaggregated, microlevel data, observed at far greater frequency. Some important examples are stock market data for which daily returns data are routinely available, and exchange rate movements, which have been tabulated on an almost continuous basis. In these settings, analysts have found that the tools of spectral analysis, and the frequency domain, have provided many useful results and have been applied to great advantage. This section introduces a small amount of the terminology of spectral analysis to acquaint the reader with a few basic features of the technique. For those who desire further detail, Fuller (1976), Granger and Newbold (1996), Hamilton (1994), Chatfield (1996), Shumway (1988), and Hatanaka (1996) (among many others with direct application in economics) are excellent introductions. Most of the following is based on Chapter 6 of Hamilton (1994).

In this framework, we view an observed time series as a weighted sum of underlying series that have different cyclical patterns. For example, aggregate retail sales and construction data display several different kinds of cyclical variation, including a regular seasonal pattern and longer frequency variation associated with variation in the economy as a whole over the business cycle. The total variance of an observed time series may thus be viewed as a sum of the contributions of these underlying series, which vary

at different frequencies. The standard application we consider is how spectral analysis is used to decompose the variance of a time series.

#### 20.2.7.a. Theoretical Results

Let $\{y_t\}_{t=-\infty,\infty}$ define a zero mean, stationary time-series process. The autocovariance at lag $k$ was defined in Section 20.2.2 as

$$\lambda_k = \lambda_{-k} = \text{Cov}[y_t, y_{t-k}].$$

We assume that the series $\lambda_k$ is *absolutely summable;* $\sum_{i=0}^{\infty} |\lambda_k|$ is finite. The **autocovariance generating function** for this time-series process is

$$g_Y(z) = \sum_{k=-\infty}^{\infty} \lambda_k z^k.$$

We evaluate this function at the complex value $z = \exp(i\omega)$, where $i = \sqrt{-1}$ and $\omega$ is a real number, and divide by $2\pi$ to obtain the **spectrum,** or **spectral density function,** of the time-series process,

$$h_Y(\omega) = \frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty} \lambda_k e^{-i\omega k}\right). \tag{20-16}$$

The spectral density function is a characteristic of the time-series process very much like the sequence of autocovariances (or the sequence of moments for a probability distribution). For a time-series process that has the set of autocovariances $\lambda_k$, the spectral density can be computed at any particular value of $\omega$. Several results can be combined to simplify $h_Y(\omega)$:

1. Symmetry of the autocovariances, $\lambda_k = \lambda_{-k}$;
2. DeMoivre's theorem, $\exp(\pm i\omega k) = \cos(\omega k) \pm i \sin(\omega k)$;
3. Polar values, $\cos(0) = 1, \cos(\pi) = 0, \sin(0) = 0, \sin(\pi) = 1$;
4. Symmetries of sin and cos functions, $\sin(-\omega) = -\sin(\omega)$ and $\cos(-\omega) = \cos(\omega)$.

One of the convenient consequences of result 2 is $\exp(i\omega k) + \exp(-i\omega k) = 2\cos(\omega k)$, which is always real. These equations can be combined to simplify the spectrum.

$$h_Y(\omega) = \frac{1}{2\pi}\left[\lambda_0 + 2\sum_{k=1}^{\infty} \lambda_k \cos(\omega k)\right], \quad \omega \in [0, \pi]. \tag{20-17}$$

This is a strictly real-valued, continuous function of $\omega$. Since the cosine function is cyclic with period $2\pi$, $h_Y(\omega) = h_Y(\omega + M2\pi)$ for any integer $M$, which implies that the entire spectrum is known if its values for $\omega$ from 0 to $\pi$ are known. [Since $\cos(-\omega) = \cos(\omega)$, $h_Y(\omega) = h_Y(-\omega)$, so the values of the spectrum for $\omega$ from 0 to $-\pi$ are the same as those from 0 to $+\pi$.] There is also a correspondence between the spectrum and the autocovariances,

$$\lambda_k = \int_{-\pi}^{\pi} h_Y(\omega) \cos(k\omega)\, d\omega,$$

which we can interpret as indicating that the sequence of autocovariances and the spectral density function just produce two different ways of looking at the same

time-series process (in the first case, in the "time domain," and in the second case, in the "frequency domain," hence the name for this analysis).

The spectral density function is a function of the infinite sequence of autocovariances. For ARMA processes, however, the autocovariances are functions of the usually small numbers of parameters, so $h_Y(\omega)$ will generally simplify considerably. For the ARMA$(p, q)$ process defined in (20-6),

$$(y_t - \mu) = \gamma_1(y_{t-1} - \mu) + \cdots + \gamma_p(y_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

or

$$\Gamma(L)(y_t - \mu) = \Theta(L)\varepsilon_t,$$

the autocovariance generating function is

$$g_Y(z) = \frac{\sigma^2\Theta(z)\Theta(1/z)}{\Gamma(z)\Gamma(1/z)} = \sigma^2\Pi(z)\Pi(1/z),$$

where $\Pi(z)$ gives the sequence of coefficients in the infinite moving-average representation of the series, $\Theta(z)/\Gamma(z)$. See, for example, (201), where this result is derived for the ARMA(2, 1) process. In some cases, this result can be used explicitly to derive the spectral density function. The spectral density function can be obtained from this relationship through

$$h_Y(\omega) = \frac{\sigma^2}{2\pi}\Pi(e^{-i\omega})\Pi(e^{i\omega}).$$

***Example 20.2 Spectral Density Function for an AR(1) Process***

For an AR(1) process with autoregressive parameter $\rho$, $y_t = \rho y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N[0, 1]$, the lag polynomials are $\boldsymbol{\Theta}(z) = 1$ and $\boldsymbol{\Gamma}(z) = 1 - \rho z$. The autocovariance generating function is

$$\begin{aligned} g_Y(z) &= \frac{\sigma^2}{(1 - \rho z)(1 - \rho/z)} \\ &= \frac{\sigma^2}{1 + \rho^2 - \rho(z + 1/z)} \\ &= \frac{\sigma^2}{1 + \rho^2}\sum_{i=0}^{\infty}\left(\frac{\rho}{1 + \rho^2}\right)^i\left(\frac{1 + z^2}{z}\right)^i. \end{aligned}$$

The spectral density function is

$$h_Y(\omega) = \frac{\sigma^2}{2\pi}\frac{1}{[1 - \rho\exp(-i\omega)][1 - \rho\exp(i\omega)]} = \frac{\sigma^2}{2\pi}\frac{1}{[1 + \rho^2 - 2\rho\cos(\omega)]}.$$

For the general case suggested at the outset, $\Gamma(L)(y_t - \mu) = \Theta(L)\varepsilon_t$, there is a template we can use, which, if not simple, is at least transparent. Let $\alpha_i$ be the reciprocal of a root of the characteristic polynomial for the autoregressive part of the model, $\Gamma(\alpha_i) = 0, i = 1, \ldots, p$, and let $\delta_j$, $j = 1, \ldots, q$, be the same for the moving-average part of the model. Then

$$h_Y(\omega) = \frac{\sigma^2}{2\pi}\frac{\prod_{j=1}^{q}[1 + \delta_j^2 - 2\delta_j\cos(\omega)]}{\prod_{i=1}^{p}[1 + \alpha_i^2 - 2\alpha_i\cos(\omega)]}.$$

Some of the roots of either polynomial may be complex pairs, but in this case, the product for adjacent pairs $(a \pm bi)$ is real, so the function is always real valued. [Note also that $(a \pm bi)^{-1} = (a \mp bi)/(a^2 + b^2)$.]

For purposes of our initial objective, decomposing the variance of the time series, our final useful theoretical result is

$$\int_{-\pi}^{\pi} h_Y(\omega)\, d\omega = \lambda_0.$$

Thus, the total variance can be viewed as the sum of the spectral densities over all possible frequencies. (More precisely, it is the area under the spectral density.) Once again exploiting the symmetry of the cosine function, we can rewrite this equation in the form

$$2\int_{0}^{\pi} h_Y(\omega)\, d\omega = \lambda_0.$$

Consider, then, integration over only some of the frequencies;

$$\frac{2}{\lambda_0}\int_{0}^{\omega_j} h_Y(\omega)\, d\omega = \tau(\omega_j), \quad 0 < \omega_j \leq \pi, 0 < \tau(\omega_j) \leq 1.$$

Thus, $\tau(\omega_j)$ can be interpreted as the proportion of the total variance of the time series that is associated with frequencies less than or equal to $\omega_j$.

#### 20.2.7.b. Empirical Counterparts

We have in hand a sample of observations, $y_t, t = 1, \ldots, T$. The first task is to establish a correspondence between the frequencies $0 < \omega \leq \pi$ and something of interest in the sample. The lowest frequency we could observe would be once in the entire sample period, so we map $\omega_1$ to $2\pi/T$. The highest would then be $\omega_T = 2\pi$, and the intervening values will be $2\pi_j/T$, $j = 2, \ldots, T-1$. It may be more convenient to think in terms of period rather than frequency. The number of periods per cycle will correspond to $T/j = 2\pi/\omega_j$. Thus, the lowest frequency, $\omega_1$, corresponds to the highest period, $T$ "dates" (months, quarters, years, etc.).

There are a number of ways to estimate the population spectral density function. The obvious way is the sample counterpart to the population spectrum. The sample of $T$ observations provides the variance and $T-1$ distinct sample autocovariances

$$c_k = c_{-k} = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y}), \qquad \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t, \quad k = 0, 1, \ldots, T-1,$$

so we can compute the **sample periodogram,** which is

$$\hat{h}_Y(\omega) = \frac{1}{2\pi}\left[c_0 + 2\sum_{k=1}^{T-1} c_k \cos(\omega k)\right].$$

The sample periodogram is a natural estimator of the spectrum, but it has a statistical flaw. With the sample variance and the $T-1$ autocovariances, we are estimating $T$ parameters with $T$ observations. The periodogram is, in the end, $T$ transformations of these $T$ estimates. As such, there are no "degrees of freedom"; the estimator does not improve as the sample size increases. A number of methods have been suggested for improving the behavior of the estimator. Two common ways are truncation and

windowing [see Chatfield (1996, pp. 139–143)]. The truncated estimator of the periodogram is based on a subset of the first $L < T$ autocovariances. The choice of $L$ is a problem because there is no theoretical guidance. Chatfield (1996) suggests $L$ approximately equal to $2\sqrt{T}$ is large enough to provide resolution while removing some of the sampling variation induced by the long lags in the untruncated estimator. The second mechanism for improving the properties of the estimator is a set of weights called a **lag window.** The revised estimator is

$$\hat{h}_Y(\omega) = \frac{1}{2\pi}\left[w_0 c_0 + 2\sum_{k=1}^{L} w_k c_k \cos(\omega k)\right],$$

where the set of weights, $\{w_k, k = 0, \ldots, L\}$, is the lag window. One choice for the weights is the Bartlett window, which produces

$$\hat{h}_{Y,\text{Bartlett}}(\omega) = \frac{1}{2\pi}\left[c_0 + 2\sum_{k=1}^{L} w(k, L) c_k \cos(\omega k)\right], \qquad w(k, L) = 1 - \frac{k}{L+1}.$$

Note that this result is the same set of weights used in the Newey–West robust covariance matrix estimator in Chapter 12, with essentially the same motivation. Two others that are commonly used are the Tukey window, which has $w_k = \frac{1}{2}[1 + \cos(\pi k/L)]$, and the Parzen window, $w_k = 1 - 6[(k/L)^2 - (k/L)^3]$, if $k \leq L/2$, and $w_k = 2(1 - k/L)^3$ otherwise.

If the series has been modeled as an ARMA process, we can instead compute the fully parametric estimator based on our sample estimates of the roots of the autoregressive and moving-average polynomials. This second estimator would be

$$\hat{h}_{Y,\text{ARMA}}(\omega) = \frac{\hat{\sigma}^2}{2\pi}\frac{\prod_{j=1}^{q}\left[1 + d_j^2 - 2d_j\cos(\omega k)\right]}{\prod_{i=1}^{p}\left[1 + a_i^2 - 2a_i\cos(\omega k)\right]}.$$

Others have been suggested. [See Chatfield (1996, Chap. 7).]

Finally, with the empirical estimate of the spectrum, the variance decomposition can be approximated by summing the values around the frequencies of interest.

***Example 20.3 Spectral Analysis of the Growth Rate of Real GNP***

Appendix Table F20.2 lists quarterly observations on U.S. GNP and the implicit price deflator for GNP for 1950 through 1983. The GNP series, with its upward trend, is obviously nonstationary. We will analyze instead the quarterly growth rate, $100[\log(\text{GNP}_t/\text{price}_t) - \log(\text{GNP}_{t-1}/\text{price}_{t-1})]$. Figure 20.2 shows the resulting data. The differenced series has 135 observations.

Figure 20.3 plots the sample periodogram, with frequencies scaled so that $\omega_j = (j/T)2\pi$. The figure shows the sample periodogram for $j = 1, \ldots, 67$ (since values of the spectrum for $j = 68, \ldots, 134$) are a mirror image of the first half, we have omitted them). Figure 20.3 shows peaks at several frequencies. The effect is more easily visualized in terms of the periods of these cyclical components. The second row of labels shows the periods, computed as quarters $= T/(2j)$, where $T = 67$ quarters. There are distinct masses around 2 to 3 years that correspond roughly to the "business cycle" of this era. One might also expect seasonal effects in these quarterly data, and there are discernible spikes in the periodogram at about 0.3 year (one quarter). These spikes, however, are minor compared with the other effects in the figure. This is to be expected, because the data are seasonally adjusted already. Finally, there is a pronounced spike at about 6 years in the periodogram. The original data in Figure 20.2 do seem consistent with this result, with substantial recessions coming at intervals of 5 to 7 years from 1953 to 1980.

To underscore these results, consider what we would obtain if we analyzed the original (log) real GNP series instead of the growth rates. Figure 20.4 shows the raw data. Although there does appear to be some short-run (high-frequency) variation (around a long-run trend,

for example), the cyclical variation of this series is obviously dominated by the upward trend. If this series were viewed as a single periodic series, then we would surmise that the period of this cycle would be the entire sample interval. The frequency of the dominant part of this time series seems to be quite close to zero. The periodogram for this series, shown in Figure 20.5, is consistent with that suspicion. By far, the largest component of the spectrum is provided by frequencies close to zero.

**FIGURE 20.2** Growth Rate of U.S. Real GNP, Quarterly, 1953 to 1984.

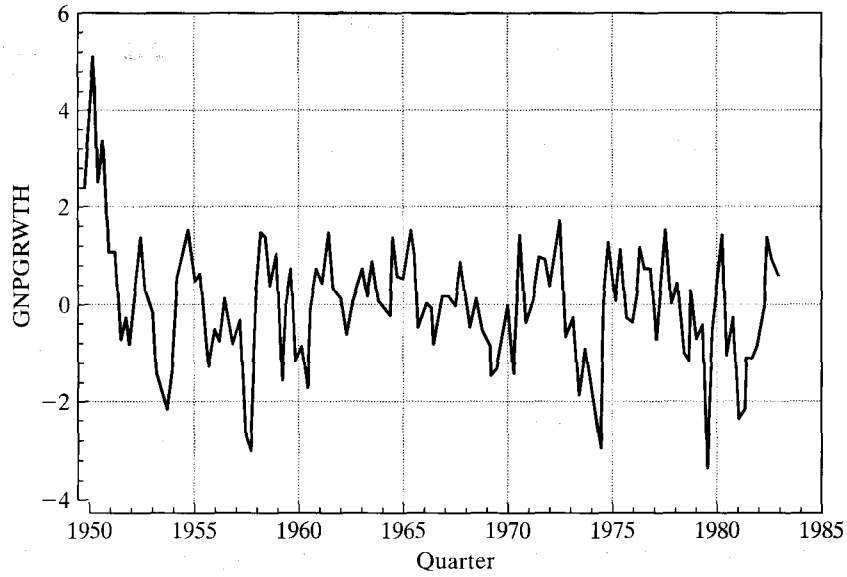
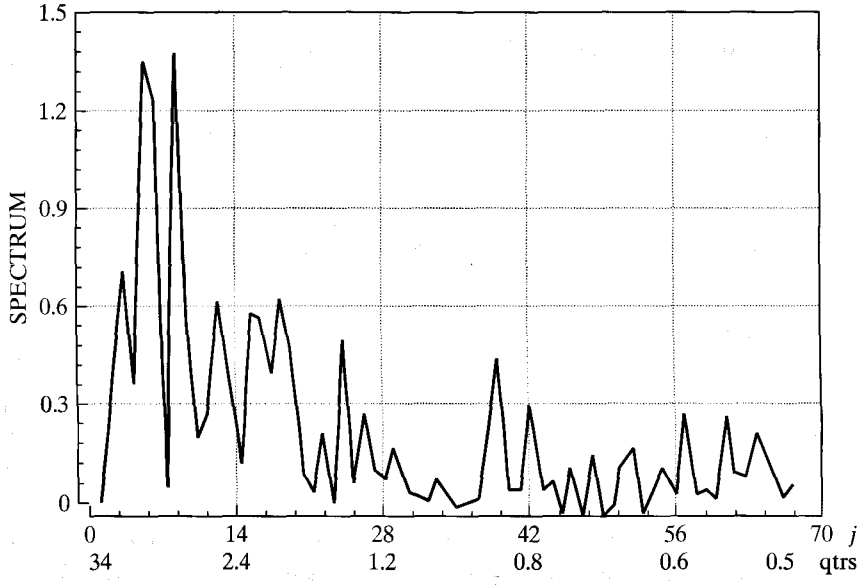



**FIGURE 20.3** Sample Periodogram.



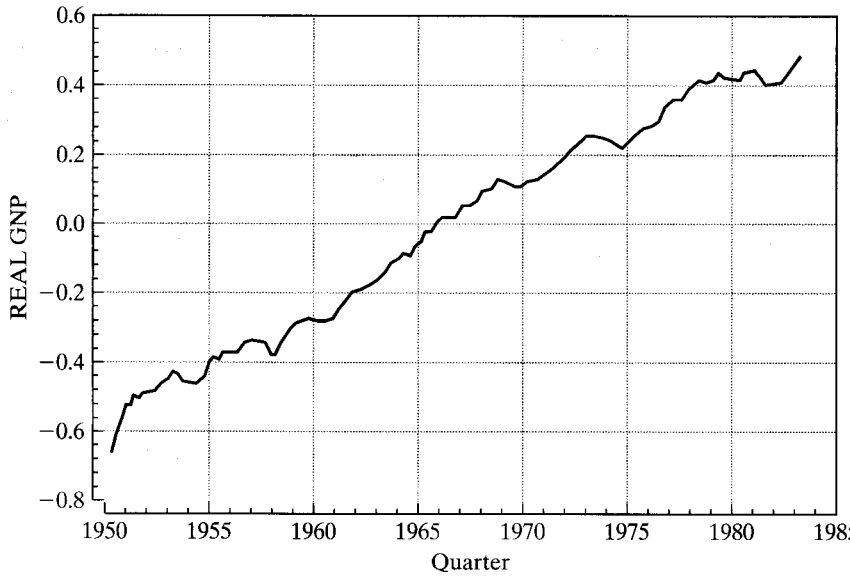



**FIGURE 20.4** Quarterly Data on Real GNP.

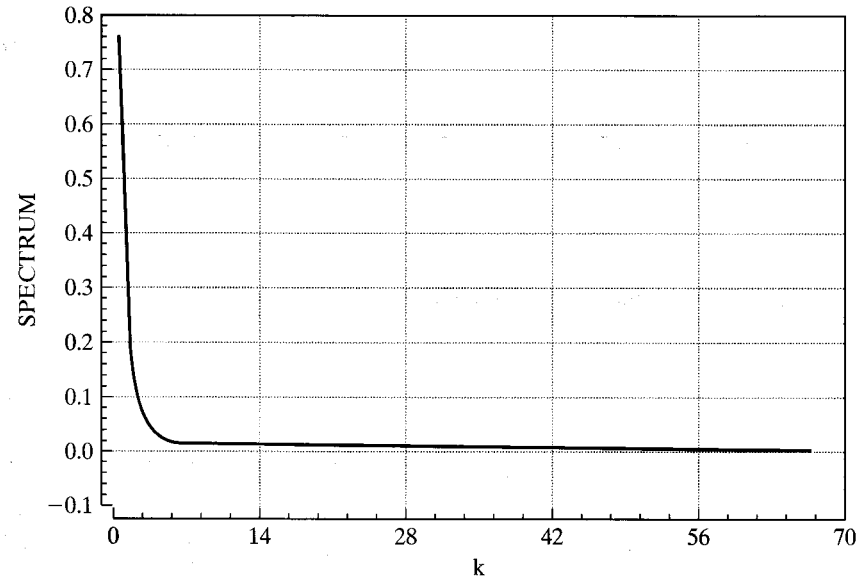



**FIGURE 20.5** Spectrum for Real GNP.

**A Computational Note** The computation in (20-16) or (20-17) is the **discrete Fourier transform** of the series of autocovariances. In principle, it involves an enormous amount of computation, on the order of $T^2$ *sets* of computations. For ordinary time series involving up to a few hundred observations, this work is not particularly onerous. (The preceding computations involving 135 observations took a total of perhaps 20 seconds of

computing time.) For series involving multiple thousands of observations, such as daily market returns, or far more, such as in recorded exchange rates and forward premiums, the amount of computation could become prohibitive. However, the computation can be done using an important tool, the fast Fourier transform (FFT), that reduces the computational level to $O(T \log_2 T)$, which is many orders of magnitude less than $T^2$. The FFT is programmed in some econometric software packages, such as RATS and Matlab. [See Press et al. (1986) for further discussion.]

## 20.3 NONSTATIONARY PROCESSES AND UNIT ROOTS

Most economic variables that exhibit strong trends, such as GDP, consumption, or the price level, are not stationary and are thus not amenable to the analysis of the previous section. In many cases, stationarity can be achieved by simple differencing or some other transformation. But, new statistical issues arise in analyzing nonstationary series that are understated by this superficial observation.

### 20.3.1 INTEGRATED PROCESSES AND DIFFERENCING

A process that figures prominently in recent work is the **random walk with drift,**

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

By direct substitution,

$$y_t = \sum_{i=0}^{\infty} (\mu + \varepsilon_{t-i}).$$

That is, $y_t$ is the simple sum of what will eventually be an infinite number of random variables, possibly with nonzero mean. If the innovations are being generated by the same zero-mean, constant-variance distribution, then the variance of $y_t$ would obviously be infinite. As such, the random walk is clearly a **nonstationary process,** even if $\mu$ equals zero. On the other hand, the first difference of $y_t$,

$$z_t = y_t - y_{t-1} = \mu + \varepsilon_t,$$

is simply the innovation plus the mean of $z_t$, which we have already assumed is stationary.

The series $y_t$ is said to be **integrated of order one,** denoted $I(1)$, because taking a first difference produces a stationary process. A nonstationary series is integrated of order $d$, denoted $I(d)$, if it becomes stationary after being first differenced $d$ times. A further generalization of the ARMA model discussed in Section 20.2.1 would be the series

$$z_t = (1 - L)^d y_t = \Delta^d y_t.$$

The resulting model is denoted an **autoregressive integrated moving-average** model, or **ARIMA** $(p, d, q)$.[14] In full, the model would be

$$\Delta^d y_t = \mu + \gamma_1 \Delta^d y_{t-1} + \gamma_2 \Delta^d y_{t-2} + \cdots + \gamma_p \Delta^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q},$$

where

$$\Delta y_t = y_t - y_{t-1} = (1 - L) y_t.$$

This result may be written compactly as

$$C(L)[(1 - L)^d y_t] = \mu + D(L)\varepsilon_t,$$

where $C(L)$ and $D(L)$ are the polynomials in the lag operator and $(1 - L)^d y_t = \Delta^d y_t$ is the $d$th difference of $y_t$.

An $I(1)$ series in its raw (undifferenced) form will typically be constantly growing, or wandering about with no tendency to revert to a fixed mean. Most macroeconomic flows and stocks that relate to population size, such as output or employment, are $I(1)$. An $I(2)$ series is growing at an ever-increasing rate. The price-level data in Appendix Table F20.2 and shown below appear to be $I(2)$. Series that are $I(3)$ or greater are extremely unusual, but they do exist. Among the few manifestly $I(3)$ series that could be listed, one would find, for example, the money stocks or price levels in hyperinflationary economies such as interwar Germany or Hungary after World War II.

***Example 20.4 A Nonstationary Series***

The nominal GDP and price deflator variables in Appendix Table F20.2 are strongly trended, so the mean is changing over time. Figures 20.6 through 20.8 plot the log of the GDP deflator series in Table F20.2 and its first and second differences. The original series and first differences are obviously nonstationary, but the second differencing appears to have rendered the series stationary.

The first 10 autocorrelations of the log of the GDP deflator series are shown in Table 20.3. The autocorrelations of the original series show the signature of a strongly trended, nonstationary series. The first difference also exhibits nonstationarity, because the autocorrelations are still very large after a lag of 10 periods. The second difference appears to be stationary, with mild negative autocorrelation at the first lag, but essentially none after that. Intuition might suggest that further differencing would reduce the autocorrelation further, but it would be incorrect. We leave as an exercise to show that, in fact, for values of $\gamma$ less than about 0.5, first differencing of an AR(1) process actually increases autocorrelation.

### 20.3.2 RANDOM WALKS, TRENDS, AND SPURIOUS REGRESSIONS

In a seminal paper, Granger and Newbold (1974) argued that researchers had not paid sufficient attention to the warning of very high autocorrelation in the residuals from conventional regression models. Among their conclusions were that macroeconomic data, as a rule, were integrated and that in regressions involving the levels of such data, the standard significance tests were usually misleading. The conventional $t$ and $F$ tests would tend to reject the hypothesis of no relationship when, in fact, there might be none.

[14] There are yet further refinements one might consider, such as removing seasonal effects from $z_t$ by differencing by quarter or month. See Harvey (1990) and Davidson and MacKinnon (1993). Some recent work has relaxed the assumption that $d$ is an integer. The **fractionally** integrated series, or ARFIMA has been used to model series in which the very long-run multipliers decay more slowly than would be predicted otherwise. See Section 20.3.5.

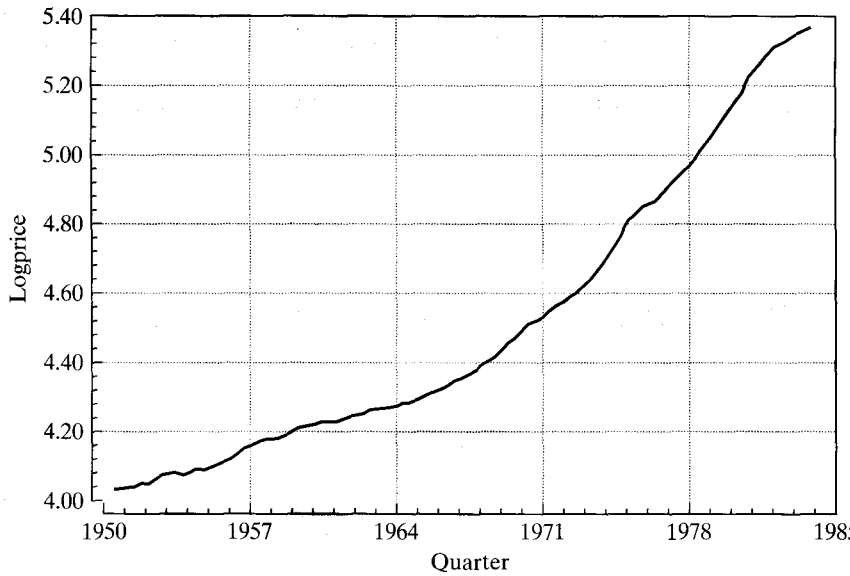



**FIGURE 20.6** Quarterly Data on log GDP Deflator.

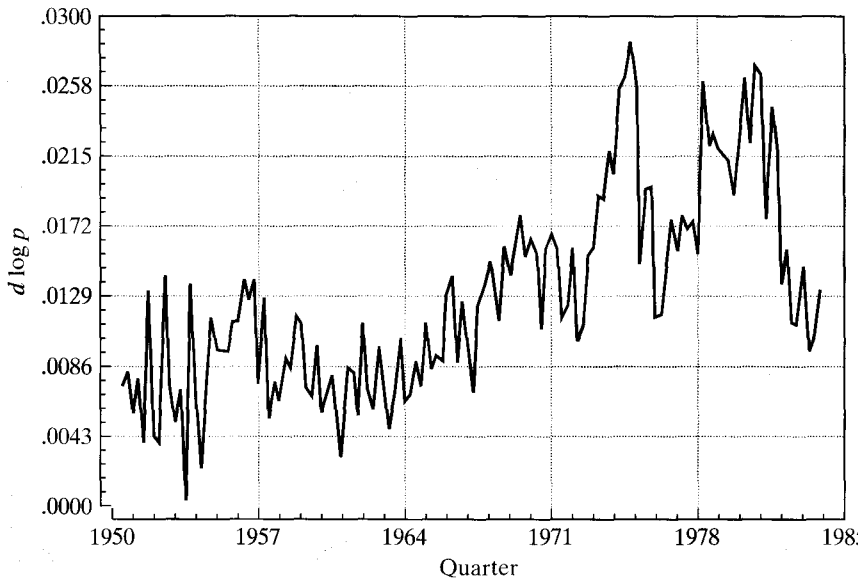



**FIGURE 20.7** First Difference of log GDP Deflator.

The general result at the center of these findings is that conventional linear regression, ignoring serial correlation, of one random walk on another is virtually certain to suggest a significant relationship, even if the two are, in fact, independent. Among their extreme conclusions, Granger and Newbold suggested that researchers use a critical $t$ value of 11.2 rather than the standard normal value of 1.96 to assess the significance of a

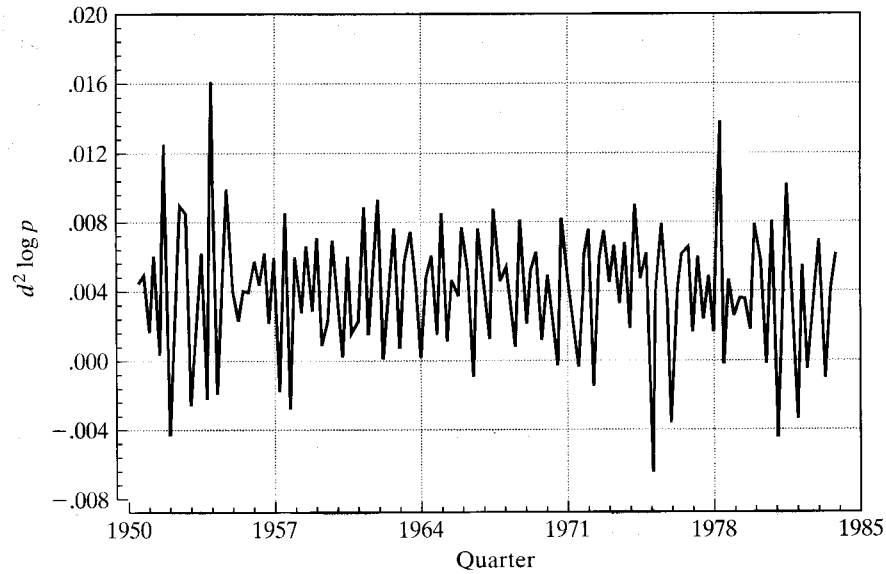



**FIGURE 20.8** Second Difference of log GNP Deflator.

**TABLE 20.3** Autocorrelations for ln GNP Deflator

| *Lag* | *Autocorrelation Function Original Series, log Price* | *Autocorrelation Function First Difference of log Price* | *Autocorrelation Function Second Difference of log Price* |
|---|---|---|---|
| 1 | 1.000 | 0.812 | −0.395 |
| 2 | 1.000 | 0.765 | −0.112 |
| 3 | 0.999 | 0.776 | 0.258 |
| 4 | 0.999 | 0.682 | −0.101 |
| 5 | 0.999 | 0.631 | −0.022 |
| 6 | 0.998 | 0.592 | 0.076 |
| 7 | 0.998 | 0.523 | −0.163 |
| 8 | 0.997 | 0.513 | 0.052 |
| 9 | 0.997 | 0.488 | −0.054 |
| 10 | 0.997 | 0.491 | 0.062 |

coefficient estimate. Phillips (1986) took strong issue with this conclusion. Based on a more general model and on an analytical rather than a Monte Carlo approach, he suggested that the normalized statistic $t_\beta/\sqrt{T}$ be used for testing purposes rather than $t_\beta$ itself. For the 50 observations used by Granger and Newbold, the appropriate critical value would be close to 15! If anything, Granger and Newbold were too optimistic.

The **random walk with drift,**

$$z_t = \mu + z_{t-1} + \varepsilon_t, \tag{20-18}$$

and the **trend stationary process,**

$$z_t = \mu + \beta t + \varepsilon_t, \tag{20-19}$$

where, in both cases, $u_t$ is a white noise process, appear to be reasonable characterizations of many macroeconomic time series.[15] Clearly both of these will produce strongly trended, nonstationary series,[16] so it is not surprising that regressions involving such variables almost always produce significant relationships. The strong correlation would seem to be a consequence of the underlying trend, whether or not there really is any regression at work. But Granger and Newbold went a step further. The intuition is less clear if there is a pure **random walk** at work,

$$z_t = z_{t-1} + \varepsilon_t, \tag{20-20}$$

but even here, they found that regression "relationships" appear to persist even in unrelated series.

Each of these three series is characterized by a **unit root.** In each case, the **data-generating process (DGP)** can be written

$$(1 - L)z_t = \alpha + \varepsilon_t, \tag{20-21}$$

where $\alpha = \mu$, $\beta$, and 0, respectively, and $v_t$ is a stationary process. Thus, the characteristic equation has a single root equal to one, hence the name. The upshot of Granger and Newbold's and Phillips's findings is that the use of data characterized by unit roots has the potential to lead to serious errors in inferences.

In all three settings, differencing or detrending would seem to be a natural first step. On the other hand, it is not going to be immediately obvious which is the correct way to proceed—the data are strongly trended in all three cases—and taking the incorrect approach will not necessarily improve matters. For example, first differencing in (20-18) or (20-20) produces a white noise series, but first differencing in (20-19) trades the trend for autocorrelation in the form of an MA(1) process. On the other hand, detrending—that is, computing the residuals from a regression on time—is obviously counterproductive in (20-18) and (20-20), even though the regression of $z_t$ on a trend will appear to be significant for the reasons we have been discussing, whereas detrending in (21-19) appears to be the right approach.[17] Since none of these approaches is likely to be obviously preferable at the outset, some means of choosing is necessary. Consider nesting all three models in a single equation,

$$z_t = \mu + \beta t + z_{t-1} + \varepsilon_t.$$

Now subtract $z_{t-1}$ from both sides of the equation and introduce the artificial parameter $\gamma$.

$$\begin{aligned} z_t - z_{t-1} &= \mu\gamma + \beta\gamma t + (\gamma - 1)z_{t-1} + \varepsilon_t \\ &= \alpha_0 + \alpha_1 t + (\gamma - 1)z_{t-1} + \varepsilon_t. \end{aligned} \tag{20-22}$$

[15]The analysis to follow has been extended to more general disturbance processes, but that complicates matters substantially. In this case, in fact, our assumption does cost considerable generality, but the extension is beyond the scope of our work. Some references on the subject are Phillips and Perron (1988) and Davidson and MacKinnon (1993).

[16]The constant term $\mu$ produces the deterministic trend in the random walk with drift. For convenience, suppose that the process starts at time zero. Then $z_t = \sum_{s=0}^{t}(\mu + \varepsilon_s) = \mu t + \sum_{s=0}^{t}\varepsilon_s$. Thus, $z_t$ consists of a deterministic trend plus a stochastic trend consisting of the sum of the innovations. The result is a variable with increasing variance around a linear trend.

[17]See Nelson and Kang (1984).

where, by hypothesis, $\gamma = 1$. Equation (20-22) provides the basis for a variety of tests for unit roots in economic data. In principle, a test of the hypothesis that $\gamma - 1$ equals zero gives confirmation of the random walk with drift, since if $\gamma$ equals 1 (and $\alpha_1$ equals zero), then (20-18) results. If $\gamma - 1$ is less than zero, then the evidence favors the trend stationary (or some other) model, and detrending (or some alternative) is the preferable approach. The practical difficulty is that standard inference procedures based on least squares and the familiar test statistics are not valid in this setting. The issue is discussed in the next section.

### 20.3.3 TESTS FOR UNIT ROOTS IN ECONOMIC DATA

The implications of unit roots in macroeconomic data are, at least potentially, profound. If a structural variable, such as real output, is truly $I(1)$, then shocks to it will have permanent effects. If confirmed, then this observation would mandate some rather serious reconsideration of the analysis of macroeconomic policy. For example, the argument that a change in monetary policy could have a transitory effect on real output would vanish.[18] The literature is not without its skeptics, however. This result rests on a razor's edge. Although the literature is thick with tests that have failed to reject the hypothesis that $\gamma = 1$, many have also not rejected the hypothesis that $\gamma \geq 0.95$, and at 0.95 (or even at 0.99), the entire issue becomes moot.[19]

Consider the simple AR(1) model with zero-mean, white noise innovations,

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

The downward bias of the least squares estimator when $\gamma$ approaches one has been widely documented.[20] For $|\gamma| < 1$, however, the least squares estimator

$$c = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$$

does have

$$\text{plim } c = \gamma$$

and

$$\sqrt{T}(c - \gamma) \xrightarrow{d} N[0, 1 - \gamma^2].$$

Does the result hold up if $\gamma = 1$? The case is called the unit root case, since in the ARMA representation $C(L)y_t = \varepsilon_t$, the characteristic equation $1 - \gamma z = 0$ has one root equal to one. That the limiting variance appears to go to zero should raise suspicions. The literature on the questions dates back to Mann and Wald (1943) and Rubin (1950). But for econometric purposes, the literature has a focal point at the celebrated papers of

[18] The 1980s saw the appearance of literally hundreds of studies, both theoretical and applied, of unit roots in economic data. An important example is the seminal paper by Nelson and Plosser (1982). There is little question but that this observation is an early part of the radical paradigm shift that has occurred in empirical macroeconomics.

[19] A large number of issues are raised in Maddala (1992, pp. 582–588).

[20] See, for example, Evans and Savin (1981, 1984).

Dickey and Fuller (1979, 1981). They showed that if $\gamma$ equals one, then

$$T(c - \gamma) \xrightarrow{d} v,$$

where $v$ is a random variable with finite, positive variance, and in finite samples, $E[c] < 1$.[21]

There are two important implications in the Dickey–Fuller results. First, the estimator of $\gamma$ is biased downward if $\gamma$ equals one. Second, the OLS estimator of $\gamma$ converges to its probability limit more rapidly than the estimators to which we are accustomed. That is, the variance of $c$ under the null hypothesis is $O(1/T^2)$, not $O(1/T)$. (In a mean squared error sense, the OLS estimator is superconsistent.) It turns out that the implications of this finding for the regressions with trended data are considerable.

We have already observed that in some cases, differencing or detrending is required to achieve stationarity of a series. Suppose, though, that the AR(1) model above is fit to an $I(1)$ series, despite that fact. The upshot of the preceding discussion is that the conventional measures will tend to hide the true value of $\gamma$; the sample estimate is biased downward, and by dint of the very small *true* sampling variance, the conventional $t$ test will tend, incorrectly, to reject the hypothesis that $\gamma = 1$. The practical solution to this problem devised by Dickey and Fuller was to derive, through Monte Carlo methods, an appropriate set of critical values for testing the hypothesis that $\gamma$ equals one in an AR(1) regression when there truly is a unit root. One of their general results is that the test may be carried out using a conventional $t$ statistic, but the critical values for the test must be revised; the standard $t$ table is inappropriate. A number of variants of this form of testing procedure have been developed. We will consider several of them.

### 20.3.4 THE DICKEY–FULLER TESTS

The simplest version of the of the model to be analyzed is the **random walk**

$$y_t = \gamma y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N[0, \sigma^2] \quad \text{and} \quad \text{Cov}[\varepsilon_t, \varepsilon_s] = 0 \, \forall \, t \neq s.$$

Under the null hypothesis that $\gamma = 1$, there are two approaches to carrying out the test. The conventional $t$ ratio

$$\text{DF}_t = \frac{\hat{\gamma} - 1}{\text{Est.Std.Error}(\hat{\gamma})}$$

with the revised set of critical values may be used for a one-sided test. Critical values for this test are shown in the top panel of Table 20.4. Note that in general, the critical value is considerably larger in absolute value than its counterpart from the $t$ distribution. The second approach is based on the statistic

$$\text{DF}_\gamma = T(\hat{\gamma} - 1).$$

Critical values for this test are shown in the top panel of Table 20.4.

The simple random walk model is inadequate for many series. Consider the rate of inflation from 1950.2 to 2000.4 (plotted in Figure 20.9) and the log of GDP over the same period (plotted in Figure 20.10). The first of these may be a random walk, but it is

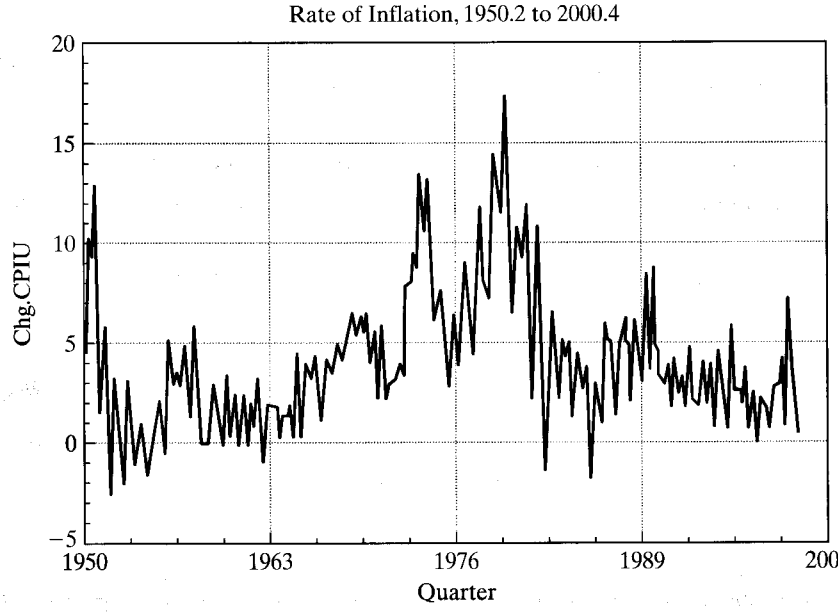
[21] A full derivation of this result is beyond the scope of this book. For the interested reader, a fairly comprehensive treatment at an accessible level is given in Chapter 17 of Hamilton (1994, pp. 475–542).

**TABLE 20.4** Critical Values for the Dickey–Fuller DF, Test

| | Sample Size | | | |
|---|---|---|---|---|
| | *25* | *50* | *100* | ∞ |
| $F$ ratio (D–F)[a] | 7.24 | 6.73 | 6.49 | 6.25 |
| $F$ ratio (standard) | 3.42 | 3.20 | 3.10 | 3.00 |
| AR model[b] (random walk) | | | | |
| 0.01 | −2.66 | −2.62 | −2.60 | −2.58 |
| 0.025 | −2.26 | −2.25 | −2.24 | −2.23 |
| 0.05 | −1.95 | −1.95 | −1.95 | −1.95 |
| 0.10 | −1.60 | −1.61 | −1.61 | −1.62 |
| 0.975 | 1.70 | 1.66 | 1.64 | 1.62 |
| AR model with constant (random walk with drift) | | | | |
| 0.01 | −3.75 | −3.59 | −3.50 | −3.42 |
| 0.025 | −3.33 | −3.23 | −3.17 | −3.12 |
| 0.05 | −2.99 | −2.93 | −2.90 | −2.86 |
| 0.10 | −2.64 | −2.60 | −2.58 | −2.57 |
| 0.975 | 0.34 | 0.29 | 0.26 | 0.23 |
| AR model with constant and time trend (trend stationary) | | | | |
| 0.01 | −4.38 | −4.15 | −4.04 | −3.96 |
| 0.025 | −3.95 | −3.80 | −3.69 | −3.66 |
| 0.05 | −3.60 | −3.50 | −3.45 | −3.41 |
| 0.10 | −3.24 | −3.18 | −3.15 | −3.13 |
| 0.975 | −0.50 | −0.58 | −0.62 | −0.66 |

[a]From Dickey and Fuller (1981, p. 1063). Degrees of freedom are 2 and $T - p - 3$.
[b]From Fuller (1976, p. 373 and 1996, Table 10.A.2).

**FIGURE 20.9** Rate of Inflation in the Consumer Price Index.



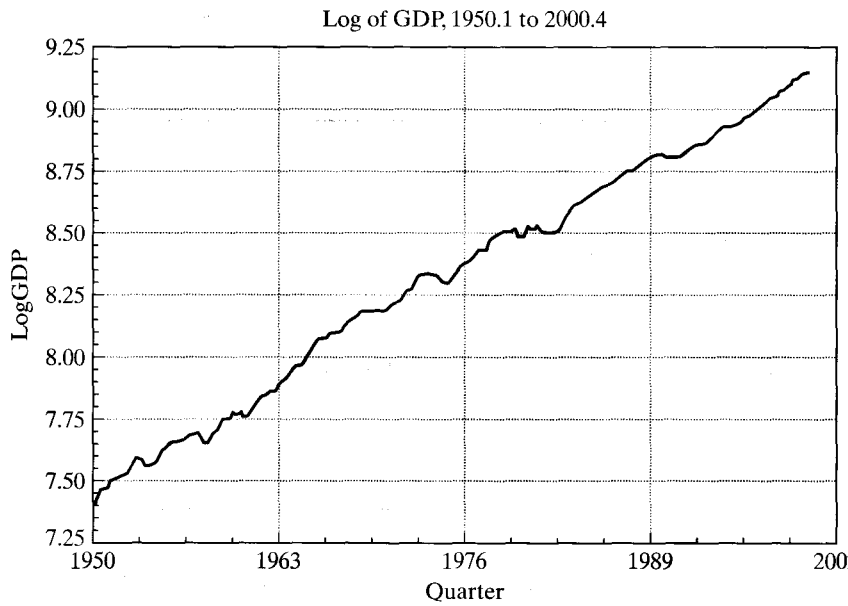



**FIGURE 20.10** Log of Gross Domestic Product.

clearly drifting. The log GDP series, in contrast, has a strong trend. For the first of these, a **random walk with drift** may be specified,

$$y_t = \mu + z_t, \qquad z_t = \gamma z_{t-1} + \varepsilon_t$$

or

$$y_t = \mu(1 - \gamma) + \gamma y_{t-1} + \varepsilon_t.$$

For the second type of series, we may specify the **trend stationary** form,

$$y_t = \mu + \beta t + z_t, \qquad z_t = \gamma z_{t-1} + \varepsilon_t$$

or

$$y_t = [\mu(1 - \gamma) + \gamma\beta] + \beta(1 - \gamma) + \gamma y_{t-1} + \varepsilon_t.$$

The tests for these forms may be carried out in the same fashion. For the model with drift only, the center panels of Tables 20.4 and 20.5 are used. When the trend is included, the lower panel of each table is used.

***Example 20.5 Tests for Unit Roots***

In Section 19.6.8, we examined Cecchetti and Rich's study of the effect of recent monetary policy on the U.S. economy. The data used in their study were the following variables:

$\pi$ = one period rate of inflation = the rate of change in the CPI
$y$ = log of real GDP
$i$ = nominal interest rate = the quarterly average yield on a 90 day T-bill
$\Delta m$ = change in the log of the money stock, M1
$i - \pi$ = ex post real interest rate
$\Delta m - \pi$ = real growth in the money stock.

**TABLE 20.5 Critical Values for the Dickey–Fuller $DF_\gamma$ Test**

| | *Sample Size* | | | |
|---|---|---|---|---|
| | ***25*** | ***50*** | ***100*** | ***∞*** |
| AR model[a] (random walk) | | | | |
| 0.01 | −11.8 | −12.8 | −13.3 | −13.8 |
| 0.025 | −9.3 | −9.9 | −10.2 | −10.5 |
| 0.05 | −7.3 | −7.7 | −7.9 | −8.1 |
| 0.10 | −5.3 | −5.5 | −5.6 | −5.7 |
| 0.975 | 1.78 | 1.69 | 1.65 | 1.60 |
| AR model with constant (random walk with drift) | | | | |
| 0.01 | −17.2 | −18.9 | −19.8 | −20.7 |
| 0.025 | −14.6 | −15.7 | −16.3 | −16.9 |
| 0.05 | −12.5 | −13.3 | −13.7 | −14.1 |
| 0.10 | −10.2 | −10.7 | −11.0 | −11.3 |
| 0.975 | 0.65 | 0.53 | 0.47 | 0.41 |
| AR model with constant and time trend (trend stationary) | | | | |
| 0.01 | −22.5 | −25.8 | −27.4 | −29.4 |
| 0.025 | −20.0 | −22.4 | −23.7 | −24.4 |
| 0.05 | −17.9 | −19.7 | −20.6 | −21.7 |
| 0.10 | −15.6 | −16.8 | −17.5 | −18.3 |
| 0.975 | −1.53 | −1.667 | −1.74 | −1.81 |

[a]From Fuller (1976, p. 373 and 1996, Table 10.A.1).

Data used in their analysis were from the period 1959.1 to 1997.4. As part of their analysis, they checked each of these series for a unit root and suggested that the hypothesis of a unit root could only be rejected for the last two variables. We will reexamine these data for the longer interval, 1950.2 to 2000.4. The data are in Appendix Table F5.1. Figures 20.11 to 20.14 show the behavior of the last four variables. The first two are shown above in Figures 20.9 and 20.10. Only the real output figure shows a strong trend, so we will use the random walk with drift for all the variables except this one.

The Dickey–Fuller tests are carried out in Table 20.6. There are 202 observations used in each one. The first observation is lost when computing the rate of inflation and the change in the money stock, and one more is lost for the difference term in the regression. The critical values from interpolating to the second row, last column in each panel for 95 percent significance and a one tailed test are −3.70 and −24.2, respectively for $DF_\tau$ and $DF_\gamma$ for the output equation, which contains the time trend and −3.14 and −16.8 for the other equations which contain a constant but no trend. For the output equation ($y$), the test statistics are

$$\mathrm{DF}_\tau = \frac{0.9584940384 - 1}{.017880922} = -2.32 > -3.44$$

and

$$\mathrm{DF}_\gamma = 202(0.9584940384 - 1) = -8.38 > -21.2.$$

Neither is less than the critical value, so we conclude (as have others) that there is a unit root in the log GDP process. The results of the other tests are shown in Table 20.6. Surprisingly, these results do differ sharply from those obtained by Cecchetti and Rich (2001) for $\pi$ and $\Delta m$. The sample period appears to matter; if we repeat the computation using Cecchetti and Rich's interval, 1959.4 to 1997.4, then $\mathrm{DF}_\tau$ equals −3.51. This is borderline, but less contradictory. For $\Delta m$ we obtain a value of −4.204 for $\mathrm{DF}_\tau$ when the sample is restricted to the shorter interval.

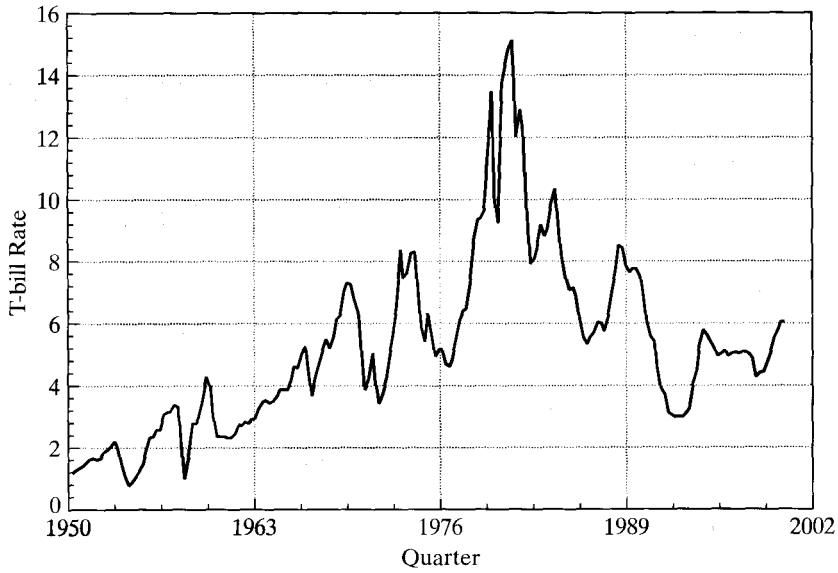



**FIGURE 20.11** T Bill Rate.

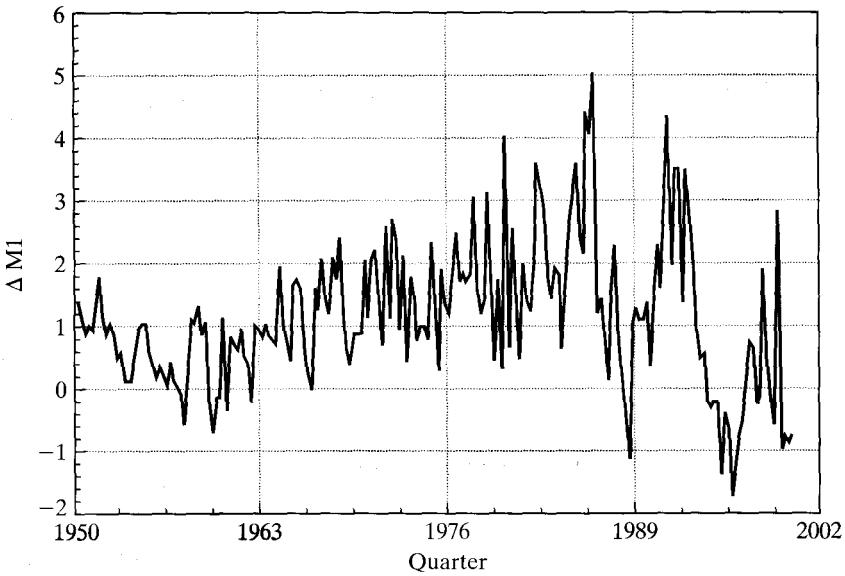



**FIGURE 20.12** Change in the Money Stock.

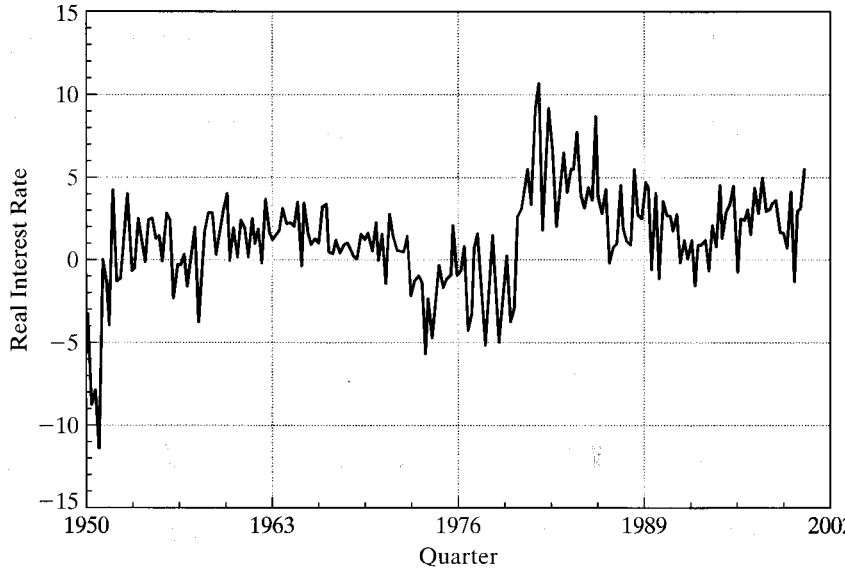



**FIGURE 20.13** Ex Post Real T Bill Rate.

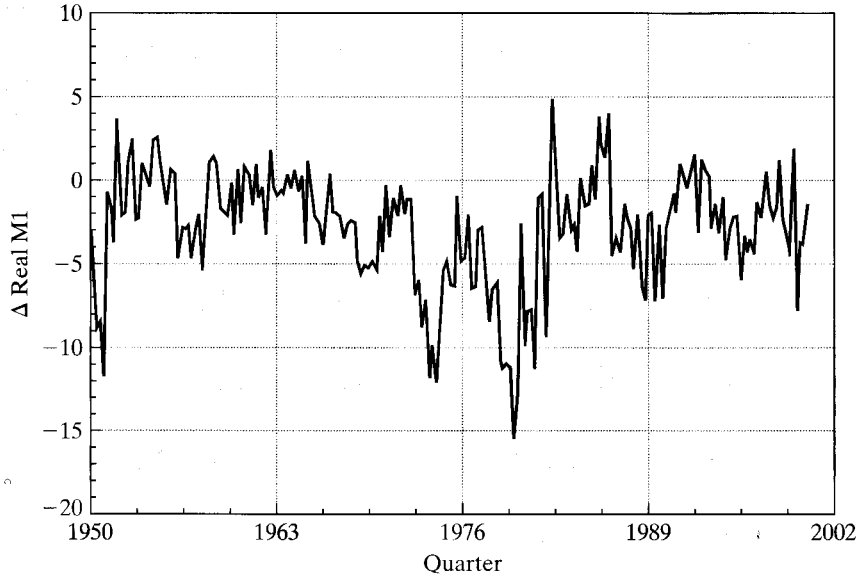



**FIGURE 20.14** Change in the Real Money Stock.

**TABLE 20.6** Unit Root Tests. (Standard errors of estimates in parentheses)

| | $\mu$ | $\beta$ | $\gamma$ | $DF_\tau$ | $DF_\gamma$ | *Conclusion* |
|---|---|---|---|---|---|---|
| $\pi$ | 0.332 | | 0.659 | −6.40 | −68.88 | Reject $H_0$ |
| | (0.0696) | | (0.0532) | $R^2 = 0.432$, $s = 0.643$ | | |
| $y$ | 0.320 | 0.00033 | 0.958 | −2.35 | −8.48 | Do not reject $H_0$ |
| | (0.134) | (0.00015) | (0.0179) | $R^2 = 0.999$, $s = 0.001$ | | |
| $i$ | 0.228 | | 0.961 | −2.14 | −7.88 | Do not reject $H_0$ |
| | (0.109) | | (0.0182) | $R^2 = 0.933$, $s = 0.743$ | | |
| $\Delta m$ | 0.448 | | 0.596 | −7.05 | −81.61 | Reject $H_0$ |
| | (0.0923) | | (0.0573) | $R^2 = 0.351$, $s = 0.929$ | | |
| $i - \pi$ | 0.615 | | 0.557 | −7.57 | −89.49 | Reject $H_0$ |
| | (0.185) | | (0.0585) | $R^2 = 0.311$, $s = 2.395$ | | |
| $\Delta m - \pi$ | 0.0700 | | 0.490 | −8.25 | −103.02 | Reject $H_0$ |
| | (0.0833) | | (0.0618) | $R^2 = 0.239$, $s = 1.176$ | | |

The Dickey–Fuller tests described above assume that the disturbances in the model as stated are white noise. An extension which will accommodate some forms of serial correlation is the **augmented Dickey–Fuller test.** The augmented Dickey–Fuller test is the same one as above, carried out in the context of the model

$$y_y = \mu + \beta t + \gamma y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + \varepsilon_t.$$

The random walk form is obtained by imposing $\mu = 0$ and $\beta = 0$; the random walk with drift has $\beta = 0$; and the trend stationary model leaves both parameters free. The two test statistics are

$$\text{DF}_\tau = \frac{\hat{\gamma} - 1}{\text{Est.Std.Error}(\hat{\gamma})}$$

exactly as constructed before and

$$\text{DF}_\gamma = \frac{T(\hat{\gamma} - 1)}{1 - \hat{\gamma}_1 - \cdots - \hat{\gamma}_p}.$$

The advantage of this formulation is that it can accommodate higher-order autoregressive processes in $\varepsilon_t$.

An alternative formulation may prove convenient. By subtracting $y_{t-1}$ from both sides of the equation, we obtain

$$\Delta y_t = \mu + \gamma^* y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \varepsilon_t,$$

where

$$\phi_j = -\sum_{k=j+1}^{p} \gamma_k \quad \text{and} \quad \gamma^* = \left(\sum_{i=1}^{p} \gamma_i\right) - 1.$$

The unit root test is carried out as before by testing the null hypothesis $\gamma^* = 0$ against $\gamma^* < 0$.[22] The $t$ test, $DF_\tau$ may be used. If the failure to reject the unit root is taken as evidence that a unit root is present, i.e., $\gamma^* = 0$, then the model specializes to the AR($p-1$) model in the first differences which is an ARIMA($p-1, 1, 0$) model for $y_t$. For a model with a time trend,

$$\Delta y_t = \mu + \beta t + \gamma^* y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \varepsilon_t,$$

the test is carried out by testing the joint hypothesis that $\beta = \gamma^* = 0$. Dickey and Fuller (1981) present counterparts to the critical $F$ statistics for testing the hypothesis. Some of their values are reproduced in the first row of Table 20.4. (Authors frequently focus on $\gamma^*$ and ignore the time trend, maintaining it only as part of the appropriate formulation. In this case, one may use the simple test of $\gamma^* = 0$ as before, with the $DF_\tau$ critical values.)

The lag length, $p$, remains to be determined. As usual, we are well advised to test down to the right value instead of up. One can take the familiar approach and sequentially examine the $t$ statistic on the last coefficient—the usual $t$ test is appropriate. An alternative is to combine a measure of model fit, such as the regression $s^2$ with one of the information criteria. The Akaike and Schwartz (Bayesian) information criteria would produce the two information measures

$$IC(p) = \ln\left(\frac{\mathbf{e}'\mathbf{e}}{T - p_{\max} - K^*}\right) + (p + K^*)\left(\frac{A^*}{T - p_{\max} - K^*}\right)$$

$K^* = 1$ for random walk, 2 for random walk with drift, 3 for trend stationary

$A^* = 2$ for Akaike criterion, $\ln(T - p_{\max} - K^*)$ for Bayesian criterion

$p_{\max}$ = the largest lag length being considered.

The remaining detail is to decide upon $p_{\max}$. The theory provides little guidance here. On the basis of a large number of simulations, Schwert (1989) found that

$$p_{\max} = \text{integer part of } [12 \times (T/100)^{.25}]$$

gave good results.

Many alternatives to the Dickey–Fuller tests have been suggested, in some cases to improve on the finite sample properties and in others to accommodate more general modeling frameworks. The Phillips (1987) and **Phillips and Perron** (1988) statistic may be computed for the same three functional forms,

$$y_y = \delta_t + \gamma y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + \varepsilon_t \qquad \textbf{(20-23)}$$

where $\delta_t$ may be 0, $\mu$, or $\mu + \beta t$. The procedure modifies the two Dickey–Fuller statistics we examined above;

$$Z_\tau = \sqrt{\frac{c_0}{a}}\left(\frac{\hat{\gamma} - 1}{v}\right) - \frac{1}{2}(a - c_0)\frac{Tv}{\sqrt{as^2}}$$

$$Z_\gamma = \frac{T(\hat{\gamma} - 1)}{1 - \hat{\gamma}_1 - \cdots - \hat{\gamma}_p} - \frac{1}{2}\left(\frac{T^2 v^2}{s^2}\right)(a - c_0)$$

[22] It is easily verified that one of the roots of the characteristic polynomial is $1/(\gamma_1 + \gamma_2 + \cdots + \gamma_p)$.

where

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-K}$$

$$v^2 = \text{estimated asymptotic variance of } \hat{\gamma}$$

$$c_j = \frac{1}{T}\sum_{s=j+1}^{T} e_t e_{t-s}, \quad j = 0, \ldots, p = j\text{th autocovariance of residuals}$$

$$c_0 = [(T-K)/T]s^2$$

$$a = c_0 + 2\sum_{j=1}^{L}\left(1 - \frac{j}{L+1}\right)c_j.$$

(Note the Newey–West (Bartlett) weights in the computation of $a$. As before, the analyst must choose $L$.) The test statistics are referred to the same Dickey–Fuller tables we have used before.

Elliot, Rothenberg, and Stock (1996) have proposed a method they denote the ADF-GLS procedure which is designed to accommodate more general formulations of $\varepsilon$; the process generating $\varepsilon_t$ is assumed to be an $I(0)$ stationary process, possibly an ARMA$(r, s)$. The null hypothesis, as before, is $\gamma = 1$ in (20-23) where $\delta_t = \mu$ or $\mu + \beta t$. The method proceeds as follows:

**Step 1.** Linearly regress

$$\mathbf{y}^* = \begin{bmatrix} y_1 \\ y_2 - \bar{r}y_1 \\ \cdots \\ y_T - \bar{r}y_{T-1} \end{bmatrix} \quad \text{on} \quad \mathbf{X}^* = \begin{bmatrix} 1 \\ 1-\bar{r} \\ \cdots \\ 1-\bar{r} \end{bmatrix} \quad \text{or} \quad \mathbf{X}^* = \begin{bmatrix} 1 & 1 \\ 1-\bar{r} & 2-\bar{r} \\ \cdots & \\ 1-\bar{r} & T-\bar{r}(T-1) \end{bmatrix}$$

for the random walk with drift and trend stationary cases, respectively. (Note that the second column of the matrix is simply $\bar{r} + (1-\bar{r})t$.) Compute the residuals from this regression, $\tilde{y}_t = y_t - \hat{\delta}_t$. $\bar{r} = 1 - 7/T$ for the random walk model and $1 - 13.5/T$ for the model with a trend.

**Step 2.** The Dickey–Fuller $DF_\tau$ test can now be carried out using the model

$$\tilde{y}_y = \gamma\tilde{y}_{t-1} + \gamma_1\Delta\tilde{y}_{t-1} + \cdots + \gamma_p\Delta\tilde{y}_{t-p} + \eta_t.$$

If the model does not contain the time trend, then the $t$ statistic for $(\gamma - 1)$ may be referred to the critical values in the center panel of Table 20.4. For the trend stationary model, the critical values are given in a table presented in Elliot et al. The 97.5 percent critical values for a one-tailed test from their table is –3.15.

As in many such cases of a new technique, as researchers develop large and small modifications of these tests, the practitioner is likely to have some difficulty deciding how to proceed. The Dickey–Fuller procedures have stood the test of time as robust tools that appear to give good results over a wide range of applications. The Phillips–Perron tests are very general, but appear to have less than optimal small sample properties. Researchers continue to examine it and the others such as Elliot et al. method. Other tests are catalogued in Maddala and Kim (1998).

***Example 20.6 Augmented Dickey–Fuller Test for a Unit Root in GDP***

The Dickey–Fuller 1981 JASA paper is a classic in the econometrics literature—it is probably the single most frequently cited paper in the field. It seems appropriate, therefore, to revisit at least some of their work. Dickey and Fuller apply their methodology to a model for the log of a quarterly series on output, the Federal Reserve Board Production Index. The model used is

$$y_t = \mu + \beta t + \gamma y_{t-1} + \phi(y_{t-1} - y_{t-2}) + \varepsilon_t. \tag{20-24}$$

The test is carried out by testing the joint hypothesis that both $\beta$ and $\gamma^*$ are zero in the model

$$y_t - y_{t-1} = \mu^* + \beta t + \gamma^* y_{t-1} + \phi(y_{t-1} - y_{t-2}) + \varepsilon_t.$$

(If $\gamma = 0$, then $\mu^*$ will also by construction.) We will repeat the study with our data on real GNP from Appendix Table F5.1 using observations 1950.1 to 2000.4.

We will use the augmented Dickey–Fuller test first. Thus, the first step is to determine the appropriate lag length for the augmented regression. Using Schwert's suggestion, we find that the maximum lag length should be allowed to reach $p_{\max} = \{\text{the integer part of } 12[204/100]^{.25}\} = 14$. The specification search uses observations 18 to 204, since as many as 17 coefficients will be estimated in the equation

$$y_t = \mu + \beta t + \gamma y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + \varepsilon_t.$$

In the sequence of 14 regressions with $j = 14, 13, \ldots$, the only statistically significant lagged difference is the first one, in the last regression, so it would appear that the model used by Dickey and Fuller would be chosen on this basis. The two information criteria produce a similar conclusion. Both of them decline monotonically from $j = 14$ all the way down to $j = 1$, so on this basis, we end the search with $j = 1$, and proceed to analyze Dickey and Fuller's model.

The linear regression results for the equation in (20-24) are

$$\begin{aligned} y_t = {} & 0.368 + 0.000391t + 0.952 y_{t-1} + 0.36025 \Delta y_{t-1} + e_t, \quad s = 0.00912 \\ & (0.125) \ (0.000138) \quad (0.0167) \quad (0.0647) \qquad\qquad R^2 = 0.999647. \end{aligned}$$

The two test statistics are

$$\text{DF}_\tau = \frac{0.95166 - 1}{0.016716} = -2.892$$

and

$$\text{DF}_c = \frac{201(0.95166 - 1)}{1 - 0.36025} = -15.263.$$

Neither statistic is less than the respective critical values, which are $-3.70$ and $-24.5$. On this basis, we conclude, as have many others, that there is a unit root in log GDP.

For the Phillips and Perron statistic, we need several additional intermediate statistics. Following Hamilton (1994, page 512), we choose $L = 4$ for the long-run variance calculation. Other values we need are $T = 201$, $\hat{\gamma} = 0.9516613$, $s^2 = 0.00008311488$, $v^2 = 0.00027942647$, and the first five autocovariances, $c_0 = 0.000081469$, $c_1 = -0.00000351162$, $c_2 = 0.00000688053$, $c_3 = 0.000000597305$, and $c_4 = -0.00000128163$. Applying these to the weighted sum produces $a = 0.0000840722$, which is only a minor correction to $c_0$. Collecting the results, we obtain the Phillips–Perron statistics, $Z_\tau = -2.89921$ and $Z_\gamma = -15.44133$. Since these are applied to the same critical values in the Dickey–Fuller tables, we reach the same conclusion as before—we do not reject the hypothesis of a unit root in log GDP.

### 20.3.5 LONG MEMORY MODELS

The autocorrelations of an integrated series [$I(1)$ or $I(2)$] display the characteristic pattern shown in Table 20.3 for the log of the GNP deflator. They remain persistently extremely high at long lags. In contrast, the autocorrelations of a stationary process typically decay at an exponential rate, so large values typically cease to appear after only a few lags. (See, e.g., the rightmost panel of Table 20.3.) Some processes appear to behave between these two benchmarks; they are clearly nonstationary, yet when differenced, they appear to show the characteristic alternating positive and negative autocorrelations, still out to long lags, that suggest "overdifferencing." But the undifferenced data show significant autocorrelations out to very long lags. Stock returns [Lo (1991)] and exchange rates [Cheung (1993)] provide some cases that have been studied. In a striking example, Ding, Granger, and Engle (1993) found significant autocorrelations out to lags of well over 2,000 days in the absolute values of daily stock market returns. [See also Granger and Ding (1996).] There is ample evidence of a lack of memory in stock market returns, but a spate of recent evidence, such as this, has been convincing that the *volatility*—the absolute value resembles the standard deviation—in stock returns has extremely long memory.

Although it is clear that an extension of the standard models of stationary time series is needed to explain the persistence of the effects of shocks on, for example, GDP and the money stock, and models of unit roots and cointegration (see Section 20.4) do appear to be helpful, there remains something of a statistical balancing act in their construction. If "the root" differs from one in either direction, then an altogether different set of statistical tools is called for. For models of very long term autocorrelation, which likewise reflect persistent response to shocks, models of long-term memory have provided a very useful extension of the concept of nonstationarity.[23] The basic building block in this class of models is the **fractionally integrated** white noise series,

$$(1 - L)^d y_t = \varepsilon_t.$$

This time series has an infinite moving-average representation if $|d| < \frac{1}{2}$, but it is nonstationary. For $d \neq 0$, the sequence of autocorrelations, $\rho_k = \lambda_k/\lambda_0$, is not absolutely summable. For this simple model,

$$\rho_k = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}.$$

The first 50 values of $\rho_k$ are shown in Figure 20.15 for $d = 0.1, 0.25, 0.4$, and 0.475. The Ding, Granger, and Engle computations display a pattern similar to that shown for 0.25 in the figure. [See Granger and Ding (1996, p. 66).] The natural extension of the model that allows for more intricate patterns in the data is the *autoregressive, fractionally integrated, moving-average,* or ARFIMA($p, d, q$) model,

$$(1 - L)^d[y_t - \gamma_1 y_{t-1} - \cdots - \gamma_p y_{t-p}] = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}, \quad y_t = Y_t - \mu.$$

[23]These models appear to have originated in the physical sciences early in the 1950s, especially with Hurst (1951), whose name is given to the effect of very long term autocorrelation in observed time series. The pioneering work in econometrics is that of Taqqu (1975), Granger and Joyeux (1980), Granger (1981), Hosking (1981), and Geweke and Porter-Hudak (1983). An extremely thorough summary and an extensive bibliography are given in Baillie (1996).

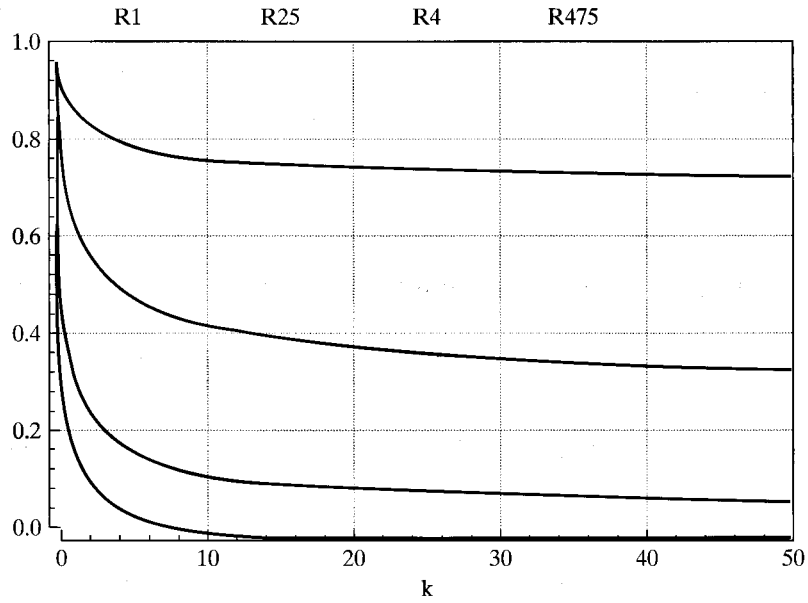



**FIGURE 20.15** Autocorrelations for a Fractionally Integrated Time Series.

Estimation of ARFIMA models is discussed in Baillie (1996) and the references cited there. A test for fractional integration effects is suggested by Geweke and Porter-Hudak (1983). The test is based on the slope in the linear regression of the logs of the first $n(T)$ values from the sample periodogram of $y_t$, that is, $z_k = \log h_Y(\omega_k)$, on the corresponding functions of the first $n(T)$ frequencies, $x_k = \log\{4\sin^2(\omega_k/2)\}$. Here $n(T)$ is taken to be reasonably small; Geweke and Porter-Hudak suggest $n(T) = \sqrt{T}$. A conventional $t$ test of the hypothesis that the slope equals zero is used to test the hypothesis.

***Example 20.7 Long-Term Memory in the Growth of Real GNP***

For the real GDP series analyzed in Example 20.6, we analyze the subseries 1950.3 to 1983.4, with $T = 135$, so we take $n(T) = 12$. The frequencies used for the periodogram are $2\pi k/135, k = 1, \ldots, 12$. The first 12 values from the periodogram are [0.05104, 0.4322, 0.7227, 0.3659, 1.353, 1.257, 0.05533, 1.388, 0.5955, 0.2043, 0.3040, 0.6381]. The linear regression produces an estimate of $d$ of 0.2505 with a standard error of 0.225. Thus, the hypothesis that $d$ equals zero cannot be rejected. This result is not surprising; the first seven autocorrelations of the series are 0.491, 0.281, 0.044, −0.076, −0.120, −0.052, and 0.018. They are trivial thereafter, suggesting that the series is, in fact, stationary. This assumption, in itself, creates something of an ambiguity. The log of the real GNP series does indeed appear to be $I(1)$. But the price level used to compute real GNP is fairly convincingly $I(2)$, or at least $I(1+d)$ for some $d$ greater than zero, judging from Figure 20.7. As such, the log of real GNP is the log of a variable that is probably at least $I(1+d)$. Although received results are not definitive, this result is probably not $I(1)$.

Models of long-term memory have been extended in many directions, and the results have been fully integrated with the unit root platform discussed earlier. Baillie's survey details many of the recently developed methods.

***Example 20.8 Long-Term Memory in Foreign Exchange Markets***

Cheung (1993) applied the long-term memory model to a study of end of week exchange rates for 16 years, 1974 to 1989. The time-series studied were the dollar spot rates of the British pound (BP), Deutsche mark (DM), Swiss franc (SF), French franc (FF), and Japanese yen (JY). Testing and estimation were done using the 1974 to 1987 data. The final 2 years of the sample were held out for out of sample forecasting.

Data were analyzed in the form of first differences of the logs so that observations are week-to-week percentage changes. Plots of the data did not suggest any obvious deviation from stationarity. As an initial assessment, the undifferenced data were subjected to augmented Dickey–Fuller tests for unit roots and the hypothesis could not be rejected. Thus, analysis proceeded using the first differences of the logs. The GPH test using $n(T) = \sqrt{T}$ for long memory in the first differences produced the following estimates of $d$, with estimated "$p$ values" in parentheses. (The $p$ value is the standard normal probability that $N[0, 1]$ is greater than or equal to the ratio of the estimated $d$ to its estimated standard error. These tests are one-sided tests. Values less than 0.05 indicate statistical significance by the usual conventions.)

| Currency | BP | DM | SF | JY | FF |
|---|---|---|---|---|---|
| $d$ | 0.1869 | 0.2943 | 0.2870 | 0.2907 | 0.4240 |
| $p$ value | (0.106) | (0.025) | (0.028) | (0.026) | (0.003) |

The unit root hypothesis is rejected in favor of the long memory model in four of the five cases. The author proceeded to estimate ARFIMA($p$, $d$, $q$) models. The coefficients of the ARFIMA models ($d$ is recomputed) are small in all cases save for the French franc, for which the estimated model is

$$(1 - L)^{0.3664}[(FF_t - \overline{FF}) - 0.4776(FF_{t-1} - \overline{FF}) - 0.1227(FF_{t-2} - \overline{FF})] = e_t + 0.8657e_{t-1}.$$

## 20.4 COINTEGRATION

Studies in empirical macroeconomics almost always involve nonstationary and trending variables, such as income, consumption, money demand, the price level, trade flows, and exchange rates. Accumulated wisdom and the results of the previous sections suggest that the appropriate way to manipulate such series is to use differencing and other transformations (such as seasonal adjustment) to reduce them to stationarity and then to analyze the resulting series as VARs or with the methods of Box and Jenkins. But recent research and a growing literature has shown that there are more interesting, appropriate ways to analyze trending variables.

In the *fully specified* regression model

$$y_t = \beta x_t + \varepsilon_t,$$

there is a presumption that the disturbances $\varepsilon_t$ are a stationary, white noise series.[24] But this presumption is unlikely to be true if $y_t$ and $x_t$ are integrated series. Generally, if two series are integrated to different orders, then linear combinations of them will be integrated to the higher of the two orders. Thus, if $y_t$ and $x_t$ are $I(1)$—that is, if both are trending variables—then we would normally expect $y_t - \beta x_t$ to be $I(1)$ regardless of the value of $\beta$, not $I(0)$ (i.e., not stationary). If $y_t$ and $x_t$ are each drifting upward

[24] If there is autocorrelation in the model, then it has been removed through an appropriate transformation.

with their own trend, then unless there is some relationship between those trends, the difference between them should also be growing, with yet another trend. There must be some kind of inconsistency in the model. On the other hand, if the two series are both $I(1)$, then there *may* be a $\beta$ such that

$$\varepsilon_t = y_t - \beta x_t$$

is $I(0)$. Intuitively, if the two series are both $I(1)$, then this partial difference between them might be stable around a fixed mean. The implication would be that the series are drifting together at roughly the same rate. Two series that satisfy this requirement are said to be **cointegrated,** and the vector $[1, -\beta]$ (or any multiple of it) is a **cointegrating vector.** In such a case, we can distinguish between a long-run relationship between $y_t$ and $x_t$, that is, the manner in which the two variables drift upward together, and the short-run dynamics, that is, the relationship between deviations of $y_t$ from its long-run trend and deviations of $x_t$ from its long-run trend. If this is the case, then differencing of the data would be counterproductive, since it would obscure the long-run relationship between $y_t$ and $x_t$. Studies of cointegration and a related technique, **error correction,** are concerned with methods of estimation that preserve the information about both forms of covariation.[25]

***Example 20.9 Cointegration in Consumption and Output***

Consumption and income provide one of the more familiar examples of the phenomenon described above. The logs of GDP and consumption for 1950.1 to 2000.4 are plotted in Figure 20.16. Both variables are obviously nonstationary. We have already verified that there is a unit root in the income data. We leave as an exercise for the reader to verify that consumption variable is likewise $I(1)$. Nonetheless, there is a clear relationship between consumption and output. To see where this discussion of relationships among variables is going, consider a simple regression of the log of consumption on the log of income, where both variables are manipulated in mean deviation form (so, the regression includes a constant). The slope in that regression is 1.056765. The residuals from the regression, $u_t = [\text{lnCons}^*, \text{lnGDP}^*][1, -1.056765]'$ (where the "*" indicates mean deviations) are plotted in Figure 20.17. The trend is clearly absent from the residuals. But, it remains to verify whether the series of residuals is stationary. In the ADF regression of the least squares residuals on a constant (random walk with drift), the lagged value and the lagged first difference, the coefficient on $u_{t-1}$ is 0.838488 (0.0370205) and that on $u_{t-1} - u_{t-2}$ is $-0.098522$. (The constant differs trivially from zero because two observations are lost in computing the ADF regression.) With 202 observations, we find $\text{DF}_\tau = -4.63$ and $\text{DF}_\gamma = -29.55$. Both are well below the critical values, which suggests that the residual series does not contain a unit root. We conclude (at least it appears so) that even after accounting for the trend, although neither of the original variables is stationary, there is a linear combination of them that is. If this conclusion holds up after a more formal treatment of the testing procedure, we will state that logGDP and log consumption are cointegrated.

***Example 20.10 Several Cointegrated Series***

The theory of purchasing power parity specifies that in long-run equilibrium, exchange rates will adjust to erase differences in purchasing power across different economies. Thus, if $p_1$ and $p_0$ are the price levels in two countries and $E$ is the exchange rate between the two currencies, then in equilibrium,

$$v_t = E_t \frac{p_{1t}}{p_{0t}} = \mu, \quad \text{a constant.}$$

[25] See, for example, Engle and Granger (1987) and the lengthy literature cited in Hamilton (1994). A survey paper on VARs and cointegration is Watson (1994).

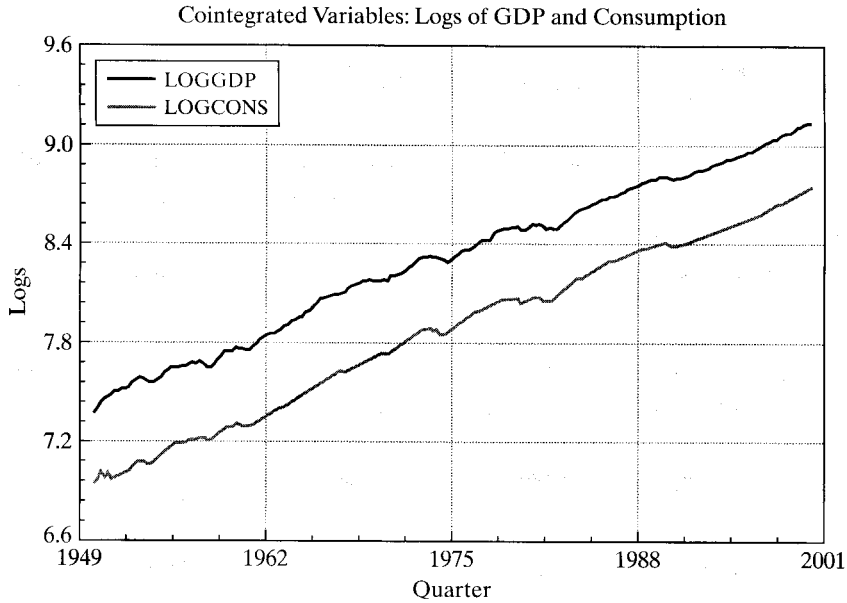



**FIGURE 20.16** Logs of Consumption and GDP.

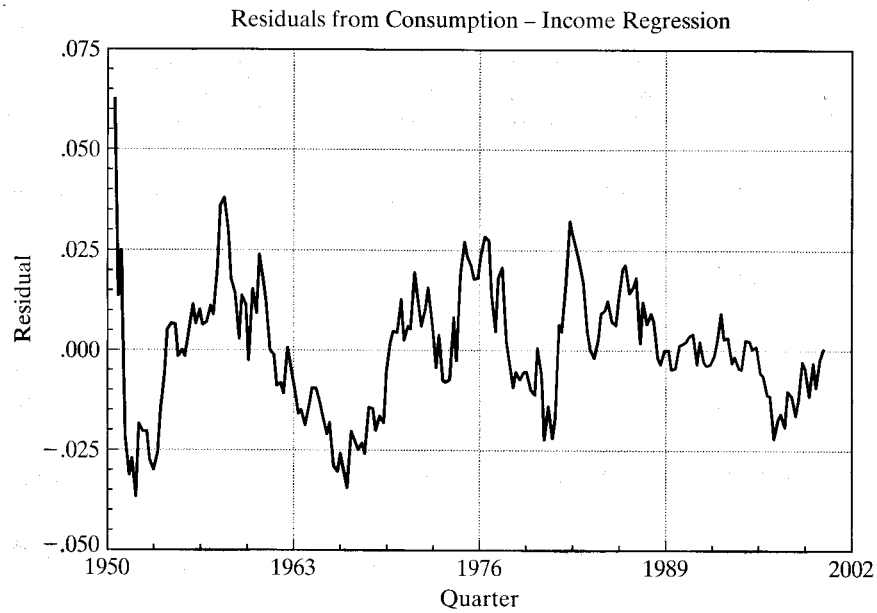



**FIGURE 20.17** Regression Residuals.

The price levels in any two countries are likely to be strongly trended. But allowing for short-term deviations from equilibrium, the theory suggests that for a particular $\beta = (\ln \mu, -1, 1)$, in the model

$$\ln E_t = \beta_1 + \beta_2 \ln p_{1t} + \beta_3 \ln p_{0t} + \varepsilon_t,$$

$\varepsilon_t = \ln v_t$ would be a stationary series, which would imply that the logs of the three variables in the model are cointegrated.

We suppose that the model involves $M$ variables, $\mathbf{y}_t = [y_{1t}, \ldots, y_{Mt}]'$, which individually may be $I(0)$ or $I(1)$, and a long-run equilibrium relationship,

$$\mathbf{y}_t'\boldsymbol{\gamma} - \mathbf{x}_t'\boldsymbol{\beta} = 0.$$

The "regressors" may include a constant, exogenous variables assumed to be $I(0)$, and/or a time trend. The vector of parameters $\boldsymbol{\gamma}$ is the cointegrating vector. In the short run, the system may deviate from its equilibrium, so the relationship is rewritten as

$$\mathbf{y}_t'\boldsymbol{\gamma} - \mathbf{x}_t'\boldsymbol{\beta} = \varepsilon_t,$$

where the **equilibrium error** $\varepsilon_t$ must be a stationary series. In fact, since there are $M$ variables in the system, at least in principle, there could be more than one cointegrating vector. In a system of $M$ variables, there can only be up to $M - 1$ linearly independent cointegrating vectors. A proof of this proposition is very simple, but useful at this point.

***Proof:*** Suppose that $\boldsymbol{\gamma}_i$ is a cointegrating vector and that there are $M$ linearly independent cointegrating vectors. Then, neglecting $\mathbf{x}_t'\boldsymbol{\beta}$ for the moment, for every $\boldsymbol{\gamma}_i$, $\mathbf{y}_t'\boldsymbol{\gamma}_i$ is a stationary series $v_{ti}$. Any linear combination of a set of stationary series is stationary, so it follows that every linear combination of the cointegrating vectors is also a cointegrating vector. If there are $M$ such $M \times 1$ linearly independent vectors, then they form a basis for the $M$-dimensional space, so any $M \times 1$ vector can be formed from these cointegrating vectors, including the columns of an $M \times M$ identity matrix. Thus, the first column of an identity matrix would be a cointegrating vector, or $y_{t1}$ is $I(0)$. This result is a contradiction, since we are allowing $y_{t1}$ to be $I(1)$. It follows that there can be at most $M - 1$ cointegrating vectors.

The number of linearly independent cointegrating vectors that exist in the equilibrium system is called its **cointegrating rank.** The cointegrating rank may range from 1 to $M - 1$. If it exceeds one, then we will encounter an interesting identification problem. As a consequence of the observation in the preceding proof, we have the unfortunate result that, in general, *if the cointegrating rank of a system exceeds one,* then without out-of-sample, *exact* information, it is not possible to estimate behavioral relationships as cointegrating vectors. Enders (1995) provides a useful example.

***Example 20.11 Multiple Cointegrating Vectors***

We consider the logs of four variables, money demand $m$, the price level $p$, real income $y$, and an interest rate $r$. The basic relationship is

$$m = \gamma_0 + \gamma_1 p + \gamma_2 y + \gamma_3 r + \varepsilon.$$

The price level and real income are assumed to be $I(1)$. The existence of long-run equilibrium in the money market implies a cointegrating vector $\boldsymbol{\alpha}_1$. If the Fed follows a certain feedback rule, increasing the money stock when *nominal* income $(y + p)$ is low and decreasing it when

nominal income is high—which might make more sense in terms of rates of growth—then there is a second cointegrating vector in which $\gamma_1 = \gamma_2$ and $\gamma_3 = 0$. Suppose that we label this vector $\alpha_2$. The parameters in the money demand equation, notably the interest elasticity, are interesting quantities, and we might seek to estimate $\alpha_1$ to learn the value of this quantity. But since every linear combination of $\alpha_1$ and $\alpha_2$ is a cointegrating vector, to this point we are only able to estimate a hash of the two cointegrating vectors.

In fact, the parameters of this model *are* identifiable from sample information (in principle). We have specified two cointegrating vectors,

$$\gamma_1 = [1, -\gamma_{10}, -\gamma_{11}, -\gamma_{12}, -\gamma_{13}]$$

and

$$\gamma_2 = [1, -\gamma_{20}, \gamma_{21}, \gamma_{21}, 0]'.$$

Although it is true that every linear combination of $\gamma_1$ and $\gamma_2$ is a cointegrating vector, only the original two vectors, as they are, have ones in the first position of both and a 0 in the last position of the second. (The equality restriction actually overidentifies the parameter matrix.) This result is, of course, exactly the sort of analysis that we used in establishing the identifiability of a simultaneous-equations system.

### 20.4.1 COMMON TRENDS

If two $I(1)$ variables are cointegrated, then some linear combination of them is $I(0)$. Intuition should suggest that the linear combination does not mysteriously create a well-behaved new variable; rather, something present in the original variables must be missing from the aggregated one. Consider an example. Suppose that two $I(1)$ variables have a linear trend,

$$y_{1t} = \alpha + \beta t + u_t,$$
$$y_{2t} = \gamma + \delta t + v_t,$$

where $u_t$ and $v_t$ are white noise. A linear combination of $y_{1t}$ and $y_{2t}$ with vector $(1, \theta)$ produces the new variable,

$$z_t = (\alpha + \theta\gamma) + (\beta + \theta\delta)t + u_t + \theta v_t,$$

which, in general, is still $I(1)$. In fact, the only way the $z_t$ series can be made stationary is if $\theta = -\beta/\delta$. If so, then the effect of combining the two variables linearly is *to remove the common linear trend,* which is the basis of Stock and Watson's (1988) analysis of the problem. But their observation goes an important step beyond this one. *The only way that $y_{1t}$ and $y_{2t}$ can be cointegrated to begin with is if they have a common trend of some sort.* To continue, suppose that instead of the linear trend $t$, the terms on the right-hand side, $y_1$ and $y_2$, are functions of a random walk, $w_t = w_{t-1} + \eta_t$, where $\eta_t$ is white noise. The analysis is identical. But now suppose that each variable $y_{it}$ has its own random walk component $w_{it}, i = 1, 2$. Any linear combination of $y_{1t}$ and $y_{2t}$ must involve *both* random walks. It is clear that they cannot be cointegrated unless, in fact, $w_{1t} = w_{2t}$. That is, once again, they must have a **common trend.** Finally, suppose that $y_{1t}$ and $y_{2t}$ share two common trends,

$$y_{1t} = \alpha + \beta t + \lambda w_t + u_t,$$
$$y_{2t} = \gamma + \delta t + \pi w_t + v_t.$$

We place no restriction on $\lambda$ and $\pi$. Then, a bit of manipulation will show that it is not possible to find a linear combination of $y_{1t}$ and $y_{2t}$ that is cointegrated, even though they share common trends. The end result for this example is that if $y_{1t}$ and $y_{2t}$ are cointegrated, then they must share exactly one common trend.

As Stock and Watson determined, the preceding is the crux of the cointegration of economic variables. A set of $M$ variables that are cointegrated can be written as a stationary component plus linear combinations of a smaller set of common trends. If the cointegrating rank of the system is $r$, then there can be up to $M - r$ linear trends and $M - r$ common random walks. [See Hamilton (1994, p. 578).] (The two-variable case is special. In a two-variable system, there can be only one common trend in total.) The effect of the cointegration is to purge these common trends from the resultant variables.

### 20.4.2 ERROR CORRECTION AND VAR REPRESENTATIONS

Suppose that the two $I(1)$ variables $y_t$ and $z_t$ are cointegrated and that the cointegrating vector is $[1, -\theta]$. Then all three variables $\Delta y_t = y_t - y_{t-1}$, $\Delta z_t$, and $(y_t - \theta z_t)$ are $I(0)$. The **error correction model**

$$\Delta y_t = \mathbf{x}_t'\boldsymbol{\beta} + \gamma(\Delta z_t) + \lambda(y_{t-1} - \theta z_{t-1}) + \varepsilon_t$$

describes the variation in $y_t$ around its long-run trend in terms of a set of $I(0)$ exogenous factors $\mathbf{x}_t$, the variation of $z_t$ around its long-run trend, and the error correction $(y_t - \theta z_t)$, which is the equilibrium error in the model of cointegration. There is a tight connection between models of cointegration and models of error correction. The model in this form is reasonable as it stands, but in fact, it is only internally consistent if the two variables are cointegrated. If not, then the third term, and hence the right-hand side, cannot be $I(0)$, even though the left-hand side must be. The upshot is that the same assumption that we make to produce the cointegration implies (and is implied by) the existence of an error correction model.[26] As we will examine in the next section, the utility of this representation is that it suggests a way to build an elaborate model of the long-run variation in $y_t$ as well as a test for cointegration. Looking ahead, the preceding suggests that residuals from an estimated cointegration model—that is, estimated equilibrium errors—can be included in an elaborate model of the long-run covariation of $y_t$ and $z_t$. Once again, we have the foundation of Engel and Granger's approach to analyzing cointegration.

Consider the VAR representation of the model

$$\mathbf{y}_t = \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where the vector $\mathbf{y}_t$ is $[y_t, z_t]'$. Now take first differences to obtain

$$\mathbf{y}_t - \mathbf{y}_{t-1} = (\boldsymbol{\Gamma} - \mathbf{I})\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

or

$$\Delta\mathbf{y}_t = \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

If all variables are $I(1)$, then all $M$ variables on the left-hand side are $I(0)$. Whether those on the right-hand side are $I(0)$ remains to be seen. The matrix $\boldsymbol{\Pi}$ produces linear

[26] The result in its general form is known as the Granger representation theorem. See Hamilton (1994, p. 582).

combinations of the variables in $\mathbf{y}_t$. But as we have seen, not all linear combinations can be cointegrated. The number of such independent linear combinations is $r < M$. Therefore, although there must be a VAR representation of the model, cointegration implies a restriction on the rank of $\mathbf{\Pi}$. It cannot have full rank; its rank is $r$. From another viewpoint, a different approach to discerning cointegration is suggested. Suppose that we estimate this model as an unrestricted VAR. The resultant coefficient matrix should be short-ranked. The implication is that if we fit the VAR model and impose short rank on the coefficient matrix as a restriction—how we could do that remains to be seen—then if the variables really are cointegrated, this restriction should not lead to a loss of fit. This implication is the basis of Johansen's (1988) and Stock and Watson's (1988) analysis of cointegration.

### 20.4.3 TESTING FOR COINTEGRATION

A natural first step in the analysis of cointegration is to establish that it is indeed a characteristic of the data. Two broad approaches for testing for cointegration have been developed. The Engle and Granger (1987) method is based on assessing whether single-equation estimates of the equilibrium errors appear to be stationary. The second approach, due to Johansen (1988, 1991) and Stock and Watson (1988), is based on the VAR approach. As noted earlier, if a set of variables is truly cointegrated, then we should be able to detect the implied restrictions in an otherwise unrestricted VAR. We will examine these two methods in turn.

Let $\mathbf{y}_t$ denote the set of $M$ variables that are believed to be cointegrated. Step one of either analysis is to establish that the variables are indeed integrated to the same order. The Dickey–Fuller tests discussed in Section 20.3.4 can be used for this purpose. If the evidence suggests that the variables are integrated to different orders or not at all, then the specification of the model should be reconsidered.

If the cointegration rank of the system is $r$, then there are $r$ independent vectors, $\boldsymbol{\gamma}_i = [1, -\boldsymbol{\theta}_i]$, where each vector is distinguished by being normalized on a different variable. If we suppose that there are also a set of $I(0)$ exogenous variables, including a constant, in the model, then each cointegrating vector produces the equilibrium relationship

$$\mathbf{y}_t'\boldsymbol{\gamma}_i = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t,$$

which we may rewrite as

$$y_{it} = \mathbf{Y}_{it}'\boldsymbol{\theta}_i + \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t.$$

We can obtain estimates of $\boldsymbol{\theta}_i$ by least squares regression. If the theory is correct *and* if this OLS estimator is consistent, then residuals from this regression should estimate the equilibrium errors. There are two obstacles to consistency. First, since both sides of the equation contain $I(1)$ variables, the problem of spurious regressions appears. Second, a moment's thought should suggest that what we have done is extract an equation from an otherwise ordinary simultaneous-equations model and propose to estimate its parameters by ordinary least squares. As we examined in Chapter 15, consistency is unlikely in that case. It is one of the extraordinary results of this body of theory that in this setting, neither of these considerations is a problem. In fact, as shown by a number of authors [see, e.g., Davidson and MacKinnon (1993)], not only is $\mathbf{c}_i$, the

OLS estimator of $\theta_i$, consistent, it is **superconsistent** in that its asymptotic variance is $O(1/T^2)$ rather than $O(1/T)$ as in the usual case. Consequently, the problem of spurious regressions disappears as well. Therefore, the next step is to estimate the cointegrating vector(s), by OLS. Under all the assumptions thus far, the residuals from these regressions, $e_{it}$, are estimates of the equilibrium errors, $\varepsilon_{it}$. As such, they should be $I(0)$. The natural approach would be to apply the familiar Dickey–Fuller tests to these residuals. The logic is sound, but the Dickey–Fuller tables are inappropriate for these estimated errors. Estimates of the appropriate critical values for the tests are given by Engle and Granger (1987), Engle and Yoo (1987), Phillips and Ouliaris (1990), and Davidson and MacKinnon (1993). If autocorrelation in the equilibrium errors is suspected, then an augmented Engle and Granger test can be based on the template

$$\Delta e_{it} = \delta e_{i,t-1} + \phi_1(\Delta e_{i,t-1}) + \cdots + u_t.$$

If the null hypothesis that $\delta = 0$ cannot be rejected (against the alternative $\delta < 0$), then we conclude that the variables are not cointegrated. (Cointegration can be rejected by this method. Failing to reject does not confirm it, of course. But having failed to reject the presence of cointegration, we will proceed as if our finding had been affirmative.)

***Example 20.9 (Continued) Consumption and Output***

In the example presented at the beginning of this discussion, we proposed precisely the sort of test suggested by Phillips and Ouliaris (1990) to determine if (log) consumption and (log) GDP are cointegrated. As noted, the logic of our approach is sound, but a few considerations remain. The Dickey–Fuller critical values suggested for the test are appropriate only in a few cases, and not when several trending variables appear in the equation. For the case of only a pair of trended variables, as we have here, one may use infinite sample values in the Dickey–Fuller tables for the trend stationary form of the equation. (The drift and trend would have been removed from the residuals by the original regression, which would have these terms either embedded in the variables or explicitly in the equation.) Finally, there remains an issue of how many lagged differences to include in the ADF regression. We have specified one, though further analysis might be called for. (A lengthy discussion of this set of issues appears in Hayashi (2000, pp. 645–648.) Thus, but for the possibility of this specification issue, the ADF approach suggested in the introduction does pass muster. The sample value found earlier was −4.63. The critical values from the table are −3.45 for 5 percent and −3.67 for 2.5 percent. Thus, we conclude (as have many other analysts) that log consumption and log GDP are cointegrated.

The Johansen (1988, 1992) and Stock and Watson (1988) methods are similar, so we will describe only the first one. The theory is beyond the scope of this text, although the operational details are suggestive. To carry out the Johansen test, we first formulate the VAR

$$\mathbf{y}_t = \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Gamma}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t.$$

The order of the model, $p$, must be determined in advance. Now, let $\mathbf{z}_t$ denote the vector of $M(p-1)$ variables,

$$\mathbf{z}_t = [\Delta\mathbf{y}_{t-1}, \Delta\mathbf{y}_{t-2}, \ldots, \Delta\mathbf{y}_{t-p+1}].$$

That is, $\mathbf{z}_t$ contains the lags 1 to $p-1$ of the first differences of all $M$ variables. Now, using the $T$ available observations, we obtain two $T \times M$ matrices of least squares residuals:

$\mathbf{D}$ = the residuals in the regressions of $\Delta\mathbf{y}_t$ on $\mathbf{z}_t$,

$\mathbf{E}$ = the residuals in the regressions of $\mathbf{y}_{t-p}$ on $\mathbf{z}_t$.

We now require the $M$ squared **canonical correlations** between the columns in $\mathbf{D}$ and those in $\mathbf{E}$. To continue, we will digress briefly to define the canonical correlations. Let $\mathbf{d}_1^*$ denote a linear combination of the columns of $\mathbf{D}$, and let $\mathbf{e}_1^*$ denote the same from $\mathbf{E}$. We wish to choose these two linear combinations so as to maximize the correlation between them. This pair of variables are the first canonical variates, and their correlation $r_1^*$ is the first canonical correlation. In the setting of cointegration, this computation has some intuitive appeal. Now, with $\mathbf{d}_1^*$ and $\mathbf{e}_1^*$ in hand, we seek a second pair of variables $\mathbf{d}_2^*$ and $\mathbf{e}_2^*$ to maximize *their* correlation, subject to the constraint that this second variable in each pair be orthogonal to the first. This procedure continues for all $M$ pairs of variables. It turns out that the computation of all these is quite simple. We will not need to compute the coefficient vectors for the linear combinations. The squared canonical correlations are simply the ordered characteristic roots of the matrix

$$\mathbf{R}^* = \mathbf{R}_{DD}^{-1/2}\mathbf{R}_{DE}\mathbf{R}_{EE}^{-1}\mathbf{R}_{ED}\mathbf{R}_{DD}^{-1/2},$$

where $\mathbf{R}_{ij}$ is the (cross-) correlation matrix between variables in set $i$ and set $j$, for $i, j = D, E$.

Finally, the null hypothesis that there are $r$ or fewer cointegrating vectors is tested using the test statistic

$$\text{TRACE TEST} = -T \sum_{i=r+1}^{M} \ln[1 - (r_i^*)^2].$$

If the correlations based on actual disturbances had been observed instead of estimated, then we would refer this statistic to the chi-squared distribution with $M - r$ degrees of freedom. Alternative sets of appropriate tables are given by Johansen and Juselius (1990) and Osterwald-Lenum (1992). Large values give evidence against the hypothesis of $r$ or fewer cointegrating vectors.

### 20.4.4 ESTIMATING COINTEGRATION RELATIONSHIPS

Both of the testing procedures discussed above involve actually estimating the cointegrating vectors, so this additional section is actually superfluous. In the Engle and Granger framework, at a second step after the cointegration test, we can use the residuals from the static regression as an error correction term in a dynamic, first-difference regression, as shown in Section 20.4.2. One can then "test down" to find a satisfactory structure. In the Johansen test shown earlier, the characteristic vectors corresponding to the canonical correlations are the sample estimates of the cointegrating vectors. Once again, computation of an error correction model based on these first step results is a natural next step. We will explore these in an application.

### 20.4.5 APPLICATION: GERMAN MONEY DEMAND

The demand for money has provided a convenient and well targeted illustration of methods of cointegration analysis. The central equation of the model is

$$m_t - p_t = \mu + \beta y_t + \gamma i_t + \varepsilon_t \qquad \textbf{(20-25)}$$

where $m_t$, $p_t$ and $y_t$ are the logs of nominal money demand, the price level and output and $i$ is the nominal interest rate (not the log of). The equation involves trending variables ($m_t$, $p_t$, $y_t$) and one which we found earlier appears to be a random walk with

drift ($i_t$). As such, the usual form of statistical inference for estimation of the income elasticity and interest semielasticity based on stationary data is likely to be misleading.

Beyer (1998) analyzed the demand for money in Germany over the period 1975 to 1994. A central focus of the study was whether the 1990 reunification produced a structural break in the long-run demand function. (The analysis extended an earlier study by the same author that was based on data which predated the reunification.) One of the interesting questions pursued in this literature concerns the stability of the long-term demand equation,

$$(m - p)_t - y_t = \mu + \gamma i_t + \varepsilon_t. \tag{20-26}$$

The left hand side is the log of the inverse of the velocity of money, as suggested by Lucas (1988). An issue to be confronted in this specification is the exogeneity of the interest variable—exogeneity [in the Engle, Hendry, and Richard (1993) sense] of income is moot in the long-run equation as its coefficient is assumed (per Lucas) to equal one. Beyer explored this latter issue in the framework developed by Engle et al. (see Section 19.6.4) and through the Granger causality testing methods discussed in Section 19.6.5.

The analytical platform of Beyer's study is a long run function for the real money stock M3 (we adopt the author's notation)

$$(m - p)^* = \delta_0 + \delta_1 y + \delta_2 RS + \delta_3 RL + \delta_4 \Delta_4 p \tag{20-27}$$

where $RS$ is a short-term interest rate, $RL$ is a long-term interest rate, and $\Delta_4 p$ is the annual inflation rate—the data are quarterly. The first step is an examination of the data. Augmented Dickey–Fuller tests suggest that for these German data in this period, $m_t$ and $p_t$ are $I(2)$ while $(m_t - p_t)$, $y_t$, $\Delta_4 p_t$, $RS_t$ and $RL_t$ are all $I(1)$. Some of Beyer's results which produced these conclusions are shown in Table 20.7. Note that though both $m_t$ and $p_t$ appear to be $I(2)$, their simple difference (linear combination) is $I(1)$, that is, integrated to a lower order. That produces the long-run specification given by (20-27). The Lucas specification is layered onto this to produce the model for the long-run velocity

$$(m - p - y)^* = \delta_0^* + \delta_2^* RS + \delta_3^* RL + \delta_4^* \Delta_4 p. \tag{20-28}$$

**TABLE 20.7** Augmented Dickey–Fuller Tests for Variables in the Beyer Model

| *Variable* | $m$ | $\Delta m$ | $\Delta^2 m$ | $p$ | $\Delta p$ | $\Delta^2 p$ | $\Delta_4 p$ | $\Delta\Delta_4 p$ |
|---|---|---|---|---|---|---|---|---|
| Spec. | TS | RW | RW | TS | RW/D | RW | RW/D | RW |
| lag | 0 | 4 | 3 | 4 | 3 | 2 | 2 | 2 |
| $DF_\tau$ | −1.82 | −1.61 | −6.87 | −2.09 | −2.14 | −10.6 | −2.66 | −5.48 |
| Crit. Value | −3.47 | −1.95 | −1.95 | −3.47 | −2.90 | −1.95 | −2.90 | −1.95 |

| *Variable* | $y$ | $\Delta y$ | $RS$ | $\Delta RS$ | $RL$ | $\Delta RL$ | $(m - p)$ | $\Delta(m - p)$ |
|---|---|---|---|---|---|---|---|---|
| Spec. | TS | RW/D | TS | RW | TS | RW | RW/D | RW/D |
| lag | 4 | 3 | 1 | 0 | 1 | 0 | 0 | 0 |
| $DF_\tau$ | −1.83 | −2.91 | −2.33 | −5.26 | −2.40 | −6.01 | −1.65 | −8.50 |
| Crit. Value | −3.47 | −2.90 | −2.90 | −1.95 | −2.90 | −1.95 | −3.47 | −2.90 |

### 20.4.5a. Cointegration Analysis and a Long Run Theoretical Model

In order for (20-27) to be a valid model, there must be at least one cointegrating vector that transforms $\mathbf{z}_t = [(m_t - p_t), y_t, RS_t, RL_t, \Delta_4 p_t]$ to stationarity. The Johansen trace test described in Section 20.4.3 was applied to the VAR consisting of these five $I(1)$ variables. A lag length of two was chosen for the analysis. The results of the trace test are a bit ambiguous; the hypothesis that $r = 0$ is rejected, albeit not strongly (sample value = 90.17 against a 95% critical value = 87.31) while the hypothesis that $r \leq 1$ is not rejected (sample value = 60.15 against a 95% critical value of 62.99). (These borderline results follow from the result that Beyer's first three eigenvalues—canonical correlations in the trace test statistic—are nearly equal. Variation in the test statistic results from variation in the correlations.) On this basis, it is concluded that the cointegrating rank equals one. The unrestricted cointegrating vector for the equation, with a time trend added is found to be

$$(m - p) = 0.936y - 1.780\Delta_4 p + 1.601RS - 3.279RL + 0.002t. \qquad \textbf{(20-29)}$$

(These are the coefficients from the first characteristic vector of the canonical correlation analysis in the Johansen computations detailed in Section 20.4.3.) An exogeneity test—we have not developed this in detail; see Beyer (1998, p. 59), Hendry and Ericsson (1991) and Engle and Hendry (1993)—confirms weak exogeneity of all four right-hand side variables in this specification. The final specification test is for the Lucas formulation and elimination of the time trend, both of which are found to pass, producing the cointegration vector

$$(m - p - y) = -1.832\Delta_4 p + 4.352RS - 10.89RL.$$

The conclusion drawn from the cointegration analysis is that a single equation model for the long run money demand is appropriate and a valid way to proceed. A last step before this analysis is a series of Granger causality tests for feedback between changes in the money stock and the four right hand variables in (20-29) (not including the trend). (See Section 19.6.5.) The test results are generally favorable, with some mixed results for exogeneity of GDP.

### 20.4.5b. Testing for Model Instability

Let $\mathbf{z}_t = [(m_t - p_t), y_t, \Delta_4 p_t, RS_t, RL_t]$ and let $\mathbf{z}_{t-1}^0$ denote the entire history of $\mathbf{z}_t$ up to the previous period. The joint distribution for $\mathbf{z}_t$, conditioned on $\mathbf{z}_{t-1}^0$ and a set of parameters $\Psi$ factors one level further into

$$f(\mathbf{z}_t \mid \mathbf{z}_{t-1}^0, \Psi) = f[(m-p)_t \mid y_t, \Delta_4 p_t, RS_t, RL_t, \mathbf{z}_{t-1}^0, \Psi_1] \times g(y_t, \Delta_4 p_t, RS_t, RL_t, \mathbf{z}_{t-1}^0, \Psi_2).$$

The result of the exogeneity tests carried out earlier implies that the conditional distribution may be analyzed apart from the marginal distribution—that is the implication of the Engle, Hendry, and Richard results noted earlier. Note the partitioning of the parameter vector. Thus, the conditional model is represented by an error correction form that explains $\Delta(m - p)_t$ in terms of its own lags, the error correction term and contemporaneous and lagged changes in the (now established) weakly exogenous

variables as well as other terms such as a constant term, trend, and certain dummy variables which pick up particular events. The error correction model specified is

$$\begin{aligned}\Delta(m-p)_t = &\sum_{i=1}^{4} c_i \Delta(m-p)_{t-i} + \sum_{i=0}^{4} d_{1,i}\Delta\left(\Delta_4 p_{t-i}\right) + \sum_{i=0}^{4} d_{2,i}\Delta y_{t-i} \\ &+ \sum_{i=0}^{4} d_{3,i}\Delta RS_{t-i} + \sum_{i=0}^{4} d_{4,i}\Delta RL_{t-i} + \lambda(m-p-y)_{t-1} \\ &+ \gamma_1 RS_{t-1} + \gamma_2 RL_{t-1} + \mathbf{d}_t'\boldsymbol{\phi} + \omega_t\end{aligned} \quad \textbf{(20-30)}$$

where $\mathbf{d}_t$ is the set of additional variables, including the constant and five one period dummy variables that single out specific events such as a currency crisis in September, 1992 [Beyer (1998, page 62, fn. 4)]. The model is estimated by least squares, "stepwise simplified and reparameterized." (The number of parameters in the equation is reduced from 32 to 15.[27])

The estimated form of (20-30) is an autoregressive distributed lag model. We proceed to use the model to solve for the long run, steady state growth path of the real money stock, (21-27). The annual growth rates $\Delta_4 m = g_m$, $\Delta_4 p = g_p$, $\Delta_4 y = g_y$ and (assumed) $\Delta_4 RS = g_{RS} = \Delta_4 RL = g_{RL} = 0$ are used for the solution

$$\frac{1}{4}(g_m - g_p) = \frac{c_4}{4}(g_m - g_p) - d_{1,1}g_p + \frac{d_{2,2}}{2}g_y + \gamma_1 RS + \gamma_2 RL + \lambda(m-p-y).^{28}$$

This equation is solved for $(m-p)^*$ under the assumption that $g_m = (g_y + g_p)$,

$$(m-p)^* = \hat{\delta}_0 + \hat{\delta}_1 g_y + y + \hat{\delta}_2 \Delta_4 p + \hat{\delta}_3 RS + \hat{\delta}_4 RL.$$

Analysis then proceeds based on this estimated long run relationship.

The primary interest of the study is the stability of the demand equation pre- and postunification. A comparison of the parameter estimates from the same set of procedures using the period 1976–1989 shows them to be surprisingly similar, $[(1.22 - 3.67g_y), 1, -3.67, 3.67, -6.44]$ for the earlier period and $[(1.25 - 2.09g_y), 1, -3.625, 3.5, -7.25]$ for the later one. This suggests, albeit informally, that the function has not changed (at least by much). A variety of testing procedures for structural break, including the Andrews and Ploberger tests discussed in Section 7.4, led to the conclusion that in spite of the dramatic changes of 1990, the long run money demand function had not materially changed in the sample period.

## 20.5 SUMMARY AND CONCLUSIONS

This chapter has completed our survey of techniques for the analysis of time-series data. While Chapter 19 was about extensions of regression modeling to time-series setting, most of the results in this Chapter focus on the internal structure of the individual time series, themselves. We began with the standard models for time-series processes. While

[27]The equation ultimately used is $\Delta(m_t - p_t) = h[\Delta(m-p)_{t-4}, \Delta\Delta_4 p_t, \Delta^2 y_{t-2}, \Delta RS_{t-1} + \Delta RS_{t-3}, \Delta^2 RL_t, RS_{t-1}, RL_{t-1}, \Delta_4 p_{t-1}, (m-p-y)_{t-1}, \mathbf{d}_t]$.

[28]The division of the coefficients is done because the intervening lags do not appear in the estimated equation.

the empirical distinction between, say AR($p$) and MA($q$) series may seem ad hoc, the Wold decomposition assures that with enough care, a variety of models can be used to analyze a time series. Section 20.2 described what is arguably the fundamental tool of modern macroeconometrics, the tests for nonstationarity. Contemporary econometric analysis of macroeconomic data has added considerable structure and formality to trending variables, which are more common than not in that setting. The variants of the Dickey–Fuller tests for unit roots are an indispensable tool for the analyst of time-series data. Section 20.4 then considered the subject of cointegration. This modeling framework is a distinct extension of the regression modeling where this discussion began. Cointegrated relationships and equilibrium relationships form the basis the time-series counterpart to regression relationships. But, in this case, it is not the conditional mean as such that is of interest. Here, both the long-run equilibrium and short-run relationships around trends are of interest and are studied in the data.

## Key Terms and Concepts

- Autoregressive integrated moving-average (ARIMA) process
- Augmented Dickey–Fuller test
- Autocorrelation
- Autocorrelation function (ACF)
- Autocovariance at lag K
- Autoregression
- Autoregressive form
- Autoregressive moving average
- Box–Jenkins analysis
- Canonical correlation
- Characteristic equation
- Cointegration
- Cointegration rank
- Cointegration relationship
- Cointegrating vector
- Common trend
- Correlogram
- Covariance stationary
- Data generating process (DGP)
- Dickey–Fuller test
- Equilibrium error
- Ergodic
- Error correction model
- Fourier transform
- Fractional integration
- Frequency domain
- Identification
- Innovation
- Integrated process
- Integrated of order one
- Invertibility
- Lag window
- Linearly deterministic component
- Linearly indeterministic component
- Moving average
- Nonstationary process
- Partial autocorrelation
- Phillips–Perron test
- Random walk
- Random walk with drift
- Sample periodogram
- Spectral density function
- Stationarity
- Square summable
- Superconsistent
- Trend stationary
- Unit root
- Univariate time series
- White noise
- Wold decomposition
- Yule–Walker equations

## Exercises

1. Find the autocorrelations and partial autocorrelations for the MA(2) process

$$\varepsilon_t = v_t - \theta_1 v_{t-1} - \theta_2 v_{t-2}.$$

2. Carry out the ADF test for a unit root in the bond yield data of Example 20.1.
3. Using the macroeconomic data in Appendix Table F5.1, estimate by least squares the parameters of the model

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \beta_3 c_{t-2} + \varepsilon_t,$$

where $c_t$ is the log of real consumption and $y_t$ is the log of real disposable income.

a. Use the Breusch and Pagan test to examine the residuals for autocorrelation.
b. Is the estimated equation stable? What is the characteristic equation for the autoregressive part of this model? What are the roots of the characteristic equation, using your estimated parameters?
c. What is your implied estimate of the short-run (impact) multiplier for change in $y_t$ on $c_t$? Compute the estimated long-run multiplier.

4. Verify the result in (20-10).
5. Show the Yule–Walker equations for an ARMA(1, 1) process.
6. Carry out an ADF test for a unit root in the rate of inflation using the subset of the data in Table F5.1 since 1974.1. (This is the first quarter after the oil shock of 1973.)
7. Estimate the parameters of the model in Example 15.1 using two-stage least squares. Obtain the residuals from the two equations. Do these residuals appear to be white noise series? Based on your findings, what do you conclude about the specification of the model?