CHAPTER 4

MODELS WITH TREND

Learning Objectives

1. Formalize simple models of variables with a time-dependent mean.
2. Compare models with deterministic versus stochastic trends.
3. Show that the so-called unit root problem arises in standard regression and in times-series models.
4. Explain how Monte Carlo and simulation techniques can be used to derive critical values for hypothesis testing.
5. Develop and illustrate the Dickey–Fuller and augmented Dickey–Fuller tests for the presence of a unit root.
6. Apply the Dickey–Fuller tests to U.S. GDP and real exchange rates.
7. Show how to apply the Dickey–Fuller test to series with serial correlation, moving average terms, multiple unit roots, and seasonal unit roots.
8. Consider tests for unit roots in the presence of structural change.
9. Illustrate the lack of power of the standard Dickey–Fuller test.
10. Show that generalized least squares (GLS) detrending methods can enhance the power of the Dickey–Fuller tests.
11. Explain how to use panel unit root tests in order to enhance the power of the Dickey–Fuller test.
12. Decompose a series with a trend into its stationary and trend components.

1. DETERMINISTIC AND STOCHASTIC TRENDS

It is helpful to represent the general solution to a linear stochastic difference equation as consisting of these three distinct parts:[1]

$$y_t = \text{trend} + \text{stationary component} + \text{noise}$$

Chapter 2 explained how to model the stationary component using the Box–Jenkins methodology. Chapter 3 showed you how to model the variance of the error (i.e., noise) component. A critical task for applied econometricians is to develop simple stochastic difference equation models that mimic the behavior of trending variables. The file RGDP.XLS contains the quarterly values of real U.S. GDP over the 1947Q1–2012Q4 period (in billions of year 2005 dollars). From the plot of the data shown in Figure 4.1, it is clear that the distinguishing feature real GDP $\{rgdp_t\}$

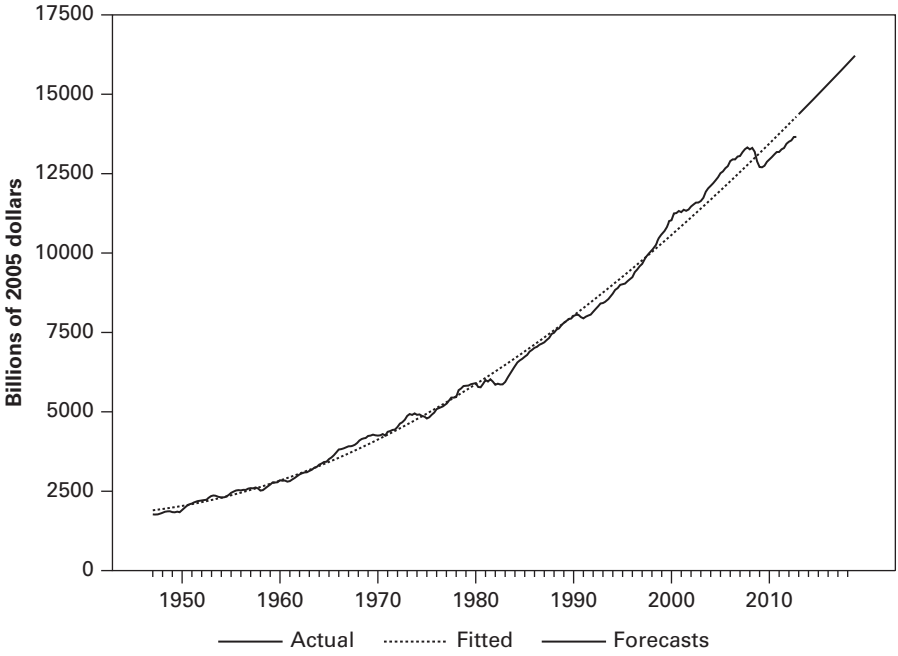


FIGURE 4.1 A Deterministic Trend in Real GDP

is that it increases over time. For such a series, a naive forecaster might estimate the sustained increase using the following cubic polynomial model for the trend:

$$\underset{}{rgdp_t} = \underset{(27.66)}{1890.247} + \underset{(4.09)}{9.108t} + \underset{(8.70)}{0.170t^2} - \underset{(-2.07)}{0.0001t^3} \tag{4.1}$$

The fitted values are shown as the dashed lines in the figure, and the forecasted values are shown as the solid line extending past 2012Q4. Regardless of the t-statistics, the use of such a model for the trend of real GDP is problematic. Since there are no stochastic components in the trend, (4.1) implies that there is a deterministic long-run growth rate of the real economy. The "Real Business Cycle" school argues that technological advancements have permanent effects on the trend of the macroeconomy. Since technological innovations are stochastic, the trend should reflect this underlying randomness. Adherents to other schools of macroeconomics would also argue that the trend is not completely deterministic. For example, they might point out that an oil price shock or a targeted tax reduction could affect investment and the economy's long-term growth rate. Moreover, the implications for the behavior of the business cycle are not credible. The deterministic trend implies that, whenever real GDP is below trend, in subsequent periods, there will be unusually high growth as real GDP returns to the trend. The reaction to the 2007–2008 financial crisis suggests that most economists and politicians do not take this notion very seriously. In fact, the forecasts beyond 2012 seem to totally ignore decline in GDP resulting from the financial crisis. Lastly, the negative coefficient on the t^3 term implies permanent declines in future GDP after a sufficiently long time.

Reexamine the 3-month T-bill rate and the yield on 5-year U.S. federal government securities shown in Figure 3.4. The two interest rates have no obvious tendency to increase or decrease. Moreover, there are no decided structural breaks that induce one-time shifts in the mean. Nevertheless, there is no pronounced tendency for either series to revert to a long-run mean. The key feature of a trend is that it has a permanent effect on a series. If the trend is defined as the permanent or nondecaying component of a time series, the two interest rates have a trend.

Suppose that a series always changes by the same fixed amount from one period to the next. To be more specific, suppose that

$$\Delta y_t = a_0$$

As you know from Chapter 1, the solution to this linear difference equation is

$$y_t = y_0 + a_0 t$$

where y_0 is the initial condition for period zero.

Hence, the solution for $\Delta y_t = a_0$ turns out to be nothing more than a deterministic linear time trend; the intercept is y_0 and the slope is a_0. Now, if we add the stationary component $A(L)\varepsilon_t$ to the trend, we obtain

$$y_t = y_0 + a_0 t + A(L)\varepsilon_t \tag{4.2}$$

In (4.2), y_t can differ from its trend value by the amount $A(L)\varepsilon_t$. Since this deviation is stationary, the $\{y_t\}$ sequence will exhibit only temporary departures from the trend. As such, the long-term forecast of y_{t+s} will converge to the trend line $y_0 + a_0(t + s)$. In the jargon of the profession, this type of model is called a **trend stationary** (TS) model.

Now suppose that the *expected* change in y_t is a_0 units. In particular, let Δy_t be equal to a_0 plus a white-noise term:

$$\Delta y_t = a_0 + \varepsilon_t \tag{4.3}$$

Sometimes, Δy_t exceeds a_0 and sometimes it falls short of a_0. Since $E_{t-1}\varepsilon_t = 0$, (4.3) implies that y_t is expected to change by a_0 units from one period to the next. The seemingly innocuous modification of (4.2) has profound differences for the trend. If y_0 is the initial condition, it is readily verified that the general solution to the first-order difference equation represented by (4.3) is

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

Here, y_t consists of the deterministic trend component $a_0 t$ and the component $y_0 + \Sigma\varepsilon_i$. We can think of this second component as a stochastic intercept term. In the absence of any shocks, the intercept is y_0. However, each ε_i shock represents a shift in the intercept. Since all values of $\{\varepsilon_i\}$ have a coefficient of unity, the effect of each shock on the intercept term is permanent. In the time-series literature, such a sequence is said to have a **stochastic trend** since each ε_i shock imparts a permanent, albeit random, change in the conditional mean of the series. If $a_0 = 0$, this type of model seems to capture some of the behavior of the interest rates. The two rates have no particular tendency to increase or decrease over time; neither do they exhibit any tendency to revert to a given mean value.

The Random Walk Model

Equation (4.3) is the basic building block for modeling series containing stochastic trends. Since these models are probably unfamiliar to you, the remainder of this section explores the nature of stochastic trends. We begin by considering the special case of (4.3) when $a_0 = 0$. This model, known as the **random walk** model, has a special place in the economics and finance literature. For example, some formulations of the efficient market hypothesis posit that the change in the price of a stock from one day to the next is completely random. As such, the current price (y_t) should be equal to last period's price plus a white-noise term, so that

$$y_t = y_{t-1} + \varepsilon_t \qquad (\text{or } \Delta y_t = \varepsilon_t)$$

Similarly, suppose you were betting on the outcome of a coin toss and a head added \$1 to your wealth while a tail cost you \$1. We could let $\varepsilon_t = +\$1$ if a head appears and $-\$1$ in the event of a tail. Thus, your current wealth (y_t) equals last period's wealth (y_{t-1}) plus the realized value of ε_t. If you play again, your wealth in $t + 1$ is $y_{t+1} = y_t + \varepsilon_{t+1}$.

If y_0 is a given initial condition, it can be readily verified that the general solution to the first-order difference equation represented by the random walk model is

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

Taking expected values, we obtain $Ey_t = Ey_{t-s} = y_0$; thus, the mean of a random walk is a constant. However, all stochastic shocks have nondecaying effects on the $\{y_t\}$ sequence. Given the first t realizations of the $\{\varepsilon_t\}$ process, the conditional mean of y_{t+1} is

$$E_t y_{t+1} = E_t(y_t + \varepsilon_{t+1}) = y_t$$

Similarly, the conditional mean of y_{t+s} (for any $s > 0$) can be obtained from

$$y_{t+s} = y_t + \sum_{i=1}^{s} \varepsilon_{t+i}$$

so that

$$E_t y_{t+s} = y_t + E_t \sum_{i=1}^{s} \varepsilon_{t+i} = y_t$$

For any positive value of s, the conditional means for all values of y_{t+s} are equivalent. Hence, the constant value of y_t is the unbiased estimator of all future values of y_{t+s}. To interpret, note that an ε_t shock has a permanent effect on y_t. This permanence is directly reflected in the forecasts for y_{t+s}.

It is easy to show that the variance is time dependent. Given the value of y_0, the variance can be constructed as

$$\text{var}(y_t) = \text{var}(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1) = t\sigma^2$$

and

$$\text{var}(y_{t-s}) = \text{var}(\varepsilon_{t-s} + \varepsilon_{t-s-1} + \cdots + \varepsilon_1) = (t-s)\sigma^2$$

Since the variance is not constant [i.e., $\text{var}(y_t) \neq \text{var}(y_{t-s})$], the random walk process is nonstationary. Moreover, as $t \to \infty$, the variance of y_t also approaches infinity. Thus, the random walk meanders without exhibiting any tendency to increase or decrease. It is also instructive to calculate the covariance of y_t and y_{t-s}. Since the mean is constant, we can form the covariance γ_{t-s} as

$$\begin{aligned} E[(y_t - y_0)(y_{t-s} - y_0)] &= E[(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)(\varepsilon_{t-s} + \varepsilon_{t-s-1} + \cdots + \varepsilon_1)] \\ &= E[(\varepsilon_{t-s})^2 + (\varepsilon_{t-s-1})^2 + \cdots + (\varepsilon_1)^2] \\ &= (t-s)\sigma^2 \end{aligned}$$

To form the correlation coefficient ρ_s, we can divide γ_{t-s} by the product of the standard deviation (SD) of y_t multiplied by the SD of y_{t-s}. Thus, the correlation coefficient ρ_s is

$$\begin{aligned} \rho_s &= (t-s)/\sqrt{(t-s)t} \\ &= [(t-s)/t]^{0.5} \end{aligned} \tag{4.4}$$

This result plays an important role in the detection of nonstationary series. For the first few autocorrelations, the sample size t will be large relative to the number of autocorrelations formed; for small values of s, the ratio $(t-s)/t$ is approximately equal to unity. However, as s increases, the values of ρ_s will decline. Hence, when using sample data, *the autocorrelation function for a random walk process will show a slight tendency to decay*. Thus, it will not be possible to use the autocorrelation function to distinguish between a unit root process and a stationary process with an autoregressive coefficient that is close to unity.

Panel (a) in Figure 4.2 shows the time path of a simulated random walk process. First, 100 normally distributed random deviates were drawn from a theoretical distribution with a mean of zero and a variance equal to unity. By setting $y_0 = 1$, each value of y_t ($t = 1, \ldots, 100$) was constructed by adding the random variable to the value of y_{t-1}. As expected, the series meanders without any tendency to revert to a long-run value. However, there does appear to be a slight positive trend in the simulated data. The reason for the upward trend is that this particular simulation happened to contain more positive values than negative values. The impression of a steadily increasing trend in the true data-generating process is false and serves as a reminder against relying solely on causal inspection.

The Random Walk Plus Drift Model

Now, let the change in y_t be partially deterministic and partially stochastic. The **random walk plus drift model** augments the random walk model by adding a constant term a_0, so that

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

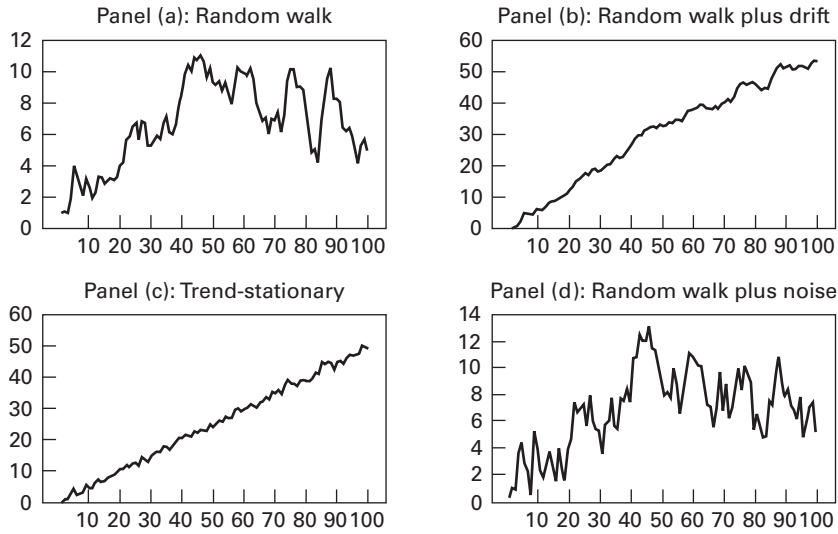


FIGURE 4.2 Four Series With Trends

Hence, you can see that (4.3) is actually a random walk plus drift process. Given the initial condition y_0, the general solution for y_t is given by

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i \tag{4.5}$$

Here, the behavior of y_t is governed by two nonstationary components: a linear deterministic trend and the stochastic trend $\Sigma\varepsilon_i$. As such, a random walk plus drift is a pure model of a trend; there is no separate stationary component in (4.5).

If we take expectations, the mean of y_t is $y_0 + a_0 t$ and the mean of y_{t+s} is $Ey_{t+s} = y_0 + a_0(t + s)$. To explain, the deterministic change in each realization of $\{y_t\}$ is a_0; after t periods, the cumulated change is $a_0 t$. In addition, there is the stochastic trend $\Sigma\varepsilon_i$; each ε_i shock has a permanent effect on the mean of y_t. Notice that the first difference of the series is stationary; taking the first difference yields the stationary sequence $\Delta y_t = a_0 + \varepsilon_t$.

Panel (b) of Figure 4.2 illustrates a simulated random walk plus drift model. The value of a_0 was set equal to 0.5, and (4.5) was simulated using the same 100 deviates used for the random walk model above. Clearly, the deterministic time trend dominates the time path of the series. In a very large sample, asymptotic theory suggests this will always be the case. However, you should not conclude that it is always easy to discern the difference between a random walk model and a model with drift. In a small sample, increasing the variance of $\{\varepsilon_t\}$ or decreasing the absolute value of a_0 could cloud the long-run properties of the sequence. Panel (c) uses the same random numbers to generate the TS series $y_t = 0.5t + \varepsilon_t$. The patterns evident in the random walk plus drift model and the TS series look strikingly similar to each other and to the real GDP series shown in Figure 4.1.

To obtain the s-step-ahead forecast for a random walk plus drift, update (4.5) by s periods to obtain

$$\begin{aligned} y_{t+s} &= y_0 + a_0(t+s) + \sum_{i=1}^{t+s} \varepsilon_i \\ &= y_t + a_0 s + \sum_{i=1}^{s} \varepsilon_{t+i} \end{aligned}$$

Taking the conditional expectation of y_{t+s}, it follows that

$$E_t y_{t+s} = y_t + a_0 s.$$

In contrast to the pure random walk model, the forecast function is not flat. The fact that the average change in y_t is always the constant a_0 is reflected in the forecast function. In addition to the given value of y_t, we project this deterministic change s times into the future.

Generalizations of the Stochastic Trend Model

It is not too difficult to generalize the random walk model to allow y_t to be the sum of a stochastic trend and a white-noise component. Formally, this third model—called a **random walk plus noise**—is represented by

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + \eta_t \tag{4.6}$$

where $\{\eta_t\}$ is a white-noise process with variance σ_η^2; and ε_t and η_{t-s} are independently distributed for all t and s [i.e., $E(\varepsilon_t \eta_{t-s}) = 0$].

If we take the first difference of (4.6), the random walk plus noise model becomes

$$\Delta y_t = \varepsilon_t + \Delta \eta_t \tag{4.7}$$

You can easily verify that (4.6) and (4.7) are equivalent by writing y_{t-1} as

$$y_{t-1} = y_0 + \sum_{i=1}^{t-1} \varepsilon_i + \eta_{t-1}$$

Subtract this expression from (4.6) to obtain (4.7). From (4.6), you can see that the key properties of the random walk plus noise model are as follows:

1. Given the value y_0, the mean of the $\{y_t\}$ sequence is constant: $Ey_t = y_0$ and updating by s periods yields $Ey_{t+s} = y_0$. Notice that the successive ε_t shocks have permanent effects on the $\{y_t\}$ sequence in that there is no decay factor on the past values of ε_i. Hence, y_t has the stochastic trend component $\Sigma\varepsilon_i$.
2. The $\{y_t\}$ sequence has a pure noise component in that the $\{\eta_t\}$ sequence has only a temporary effect on the $\{y_t\}$ sequence. The current realization of η_t affects only y_t but not the subsequent values y_{t+s}.

3. The variance of $\{y_t\}$ is not constant: $\text{var}(y_t) = t\sigma^2 + \sigma_\eta^2$ and $\text{var}(y_{t-s}) = (t-s)\sigma^2 + \sigma_\eta^2$. As in the other models with a stochastic trend, the variance of y_t approaches infinity as t increases. The presence of the noise component means that the correlation coefficient between y_t and y_{t-s} is smaller than that for the pure random walk model.

To prove that the sample correlogram will exhibit even faster decay than in the pure random walk model, note that the covariance between y_t and y_{t-s} is

$$\begin{aligned}\text{cov}(y_t, y_{t-s}) &= E[(y_t - y_0)(y_{t-s} - y_0)] \\ &= E[(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \cdots + \varepsilon_t + \eta_t)(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \cdots + \varepsilon_{t-s} + \eta_{t-s})]\end{aligned}$$

Since $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent white-noise sequences

$$\text{cov}(y_t, y_{t-s}) = (t-s)\sigma^2$$

Thus, the correlation coefficient ρ_s is

$$\rho_s = \frac{(t-s)\sigma^2}{\sqrt{(t\sigma^2 + \sigma_\eta^2)[(t-s)\sigma^2 + \sigma_\eta^2)]}}$$

Comparison of ρ_s with the correlation coefficient for the pure random walk model (i.e., equation 4.4) verifies that the autocorrelations for the random walk plus noise model are always smaller for $\sigma_\eta^2 > 0$. Panel (d) of Figure 4.2 shows a random walk plus noise model. The series was simulated by drawing a second 100 normally distributed random terms to represent the $\{\eta_t\}$ series. For each value of t, y_t was calculated using (4.6). If we compare Panels (a) and (d), it can be seen that the two series track each other quite well. The random walk plus noise model could mimic the same set of macroeconomic variables as the random walk model. The effect of the "noise" component $\{\eta_t\}$ is to increase the variance of $\{y_t\}$ without affecting its long-run behavior. After all, the random walk plus noise series is nothing more than the random walk model with a purely temporary component added.

The random walk plus noise and the random walk plus drift models are the building blocks of more complex time-series models. For example, the noise and drift components can easily be incorporated into a single model by modifying (4.7) such that the trend in y_t contains a deterministic and a stochastic component. Specifically, replace (4.7) with

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

or

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + \eta_t \tag{4.8}$$

Equation (4.8) is called the **trend plus noise model**; y_t is the sum of a deterministic trend, a stochastic trend, and a pure white-noise term. Moreover, the noise sequence does not need to be a white-noise process. Let $A(L)$ be a polynomial in the lag operator

L; it is possible to augment a random walk plus drift process with the stationary process $A(L)\eta_t$ so that the **general trend plus irregular model** is

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + A(L)\eta_t \tag{4.9}$$

Thus, (4.9) has a deterministic trend, a stochastic trend, and a stationary component.

Many more details of these **unobserved components models** are examined in Section 4.1 of the *Supplementary Manual*. It is useful to work through this section and to understand the application of **signal extraction** methods to this class of model.

2. REMOVING THE TREND

From the previous section, it should be clear that there are important differences between a series with a trend and a stationary series. Shocks to a stationary time series are necessarily temporary; over time, the effects of the shocks will dissipate, and the series will revert to its long-run mean level. On the other hand, a series containing a stochastic trend will not revert to a long-run level. Note that the trend can have deterministic and stochastic components. These components of the trend have important implications for the appropriate transformation necessary to attain a stationary series. The usual methods for eliminating the trend are **differencing** and **detrending**. For historical reasons, regressing a variable on a constant and *time* and saving the residuals is called detrending. We still use this term even though the method removes only a deterministic, not a stochastic, trend. A series containing a unit root can be made stationary by differencing. In fact, we already know that the dth difference of ARIMA(p, d, q) model is stationary. The aim of this section is to compare these two methods of isolating the trend.

Differencing

First consider the solution for the random walk plus drift model:

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

Taking the first difference, we obtain $\Delta y_t = a_0 + \varepsilon_t$. Clearly, the $\{\Delta y_t\}$ sequence—equal to a constant plus a white-noise disturbance—is stationary. Viewing Δy_t as the variable of interest, we have

$$\begin{aligned} E(\Delta y_t) &= E(a_0 + \varepsilon_t) = a_0 \\ \text{var}(\Delta y_t) &\equiv E(\Delta y_t - a_0)^2 = E(\varepsilon_t)^2 = \sigma^2 \end{aligned}$$

and for $s \neq 0$

$$\text{cov}(\Delta y_t, \Delta y_{t-s}) \equiv E[(\Delta y_t - a_0)(\Delta y_{t-s} - a_0)] = E(\varepsilon_t \varepsilon_{t-s}) = 0$$

Since the mean and variance are constants and the covariance between Δy_t and Δy_{t-s} does not depend on t, the $\{\Delta y_t\}$ sequence is stationary.

The random walk plus noise model is an interesting case study. In first differences, the model can be written as $\Delta y_t = \varepsilon_t + \Delta\eta_t$. In this form, it is easy to show that Δy_t is stationary. Clearly, the mean is zero because

$$E\Delta y_t = E(\varepsilon_t + \Delta\eta_t) = 0$$

Moreover, the variance and all autocovariances are constant and time invariant because

$$\begin{aligned}
\text{var}(\Delta y_t) &= E[(\Delta y_t)^2] = E[(\varepsilon_t + \Delta\eta_t)^2] \\
&= E[(\varepsilon_t)^2 + 2\varepsilon_t\Delta\eta_t + (\Delta\eta_t)^2] \\
&= \sigma^2 + 2E[\varepsilon_t\Delta\eta_t] + E[(\eta_t)^2 - 2\eta_t\eta_{t-1} + (\eta_{t-1})^2] = \sigma^2 + 2\sigma_\eta^2 \\
\text{cov}(\Delta y_t, \Delta y_{t-1}) &= E[(\varepsilon_t + \eta_t - \eta_{t-1})(\varepsilon_{t-1} + \eta_{t-1} - \eta_{t-2})] = -\sigma_\eta^2
\end{aligned}$$

and

$$\text{cov}(\Delta y_t, \Delta y_{t-s}) = E[(\varepsilon_t + \eta_t - \eta_{t-1})(\varepsilon_{t-s} + \eta_{t-s} - \eta_{t-s-1})] = 0 \text{ for } s > 1.$$

If we set $s = 1$, the correlation coefficient between Δy_t and Δy_{t-1} is

$$\rho_1 = \frac{\text{cov}(\Delta y_t, \Delta y_{t-1})}{\text{var}(\Delta y_t)} = \frac{-\sigma_\eta^2}{\sigma^2 + \sigma_\eta^2}$$

Examination reveals $-0.5 < \rho_1 < 0$ and that all other correlation coefficients are zero. Since the first difference of y_t acts exactly as an MA(1) process, the random walk plus noise model is ARIMA(0, 1, 1). Since adding a constant to a series has no effect on the correlogram, it additionally follows that the trend plus noise model of (4.8) also acts as an ARIMA(0,1,1) process.

Now consider the general class of ARIMA(p, d, q) models:

$$A(L)y_t = B(L)\varepsilon_t \tag{4.10}$$

where $A(L)$ and $B(L)$ are polynomials of orders p and q in the lag operator L.

First, suppose that $A(L)$ has a single unit root and that $B(L)$ has all roots outside the unit circle. We can factor $A(L)$ into two components $(1 - L)A^*(L)$, where $A^*(L)$ is a polynomial of order $p - 1$. Since $A(L)$ has only one unit root, it follows that all roots of $A^*(L)$ are outside the unit circle. Thus, we can write (4.10) as

$$(1 - L)A^*(L)y_t = B(L)\varepsilon_t$$

Now, define $y_t^* = \Delta y_t$ so that

$$A^*(L)y_t^* = B(L)\varepsilon_t \tag{4.11}$$

The $\{y_t^*\}$ sequence is stationary since all roots of $A^*(L)$ lie outside the unit circle. The point is that the first difference of a unit root process is stationary. If $A(L)$ has two unit roots, the same argument can be used to show that the second difference of $\{y_t\}$ is stationary. *The general point is that the dth difference of a process with d unit*

roots is stationary. Such a sequence is integrated of order d and is denoted by $I(d)$. An ARIMA(p, d, q) model has d unit roots; the dth difference of such a model is a stationary ARMA(p, q) process.

Detrending

We have shown that differencing can sometimes be used to transform a nonstationary model into a stationary model with an ARMA representation. This does not mean that all nonstationary models can be transformed into well-behaved ARMA models by appropriate differencing. Consider, for example, a model that is the sum of a deterministic trend and a pure noise component:

$$y_t = y_0 + a_1 t + \varepsilon_t$$

The first difference of y_t is not well-behaved because

$$\Delta y_t = a_1 + \varepsilon_t - \varepsilon_{t-1}$$

Here, Δy_t is not *invertible* in the sense that Δy_t cannot be expressed in the form of an autoregressive process. Recall that invertibility of a stationary process requires that the MA component does not have a unit root.

Instead, an appropriate way to transform this model is to estimate the regression equation $y_t = a_0 + a_1 t + \varepsilon_t$. Subtracting the estimated values of y_t from the observed series yields estimated values of the $\{\varepsilon_t\}$ series. More generally, a time series may have the polynomial trend as in

$$y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots + a_n t^n + e_t$$

where $\{e_t\}$ = a stationary process.

Detrending is accomplished by regressing $\{y_t\}$ on a deterministic polynomial time trend, as in (4.1). The appropriate degree of the polynomial can be determined by standard t-tests, F-tests, and/or using statistics such as the AIC or the SBC. The common practice is to estimate the regression equation using the largest value of n deemed reasonable. If the t-statistic indicates a_n is zero, consider a polynomial trend of order $n - 1$. Continue to pare down the order of the polynomial trend until a nonzero coefficient is found. F-tests can be used to determine whether a group of coefficients, say, a_{n-i} through a_n, is statistically different from zero. The AIC and SBC statistics can be used to reconfirm the appropriate degree of the polynomial.

Simply subtracting the estimated values of the $\{y_t\}$ sequence from the actual values yields an estimate of the stationary sequence $\{e_t\}$. The detrended process can then be modeled using traditional methods (such as ARMA estimation).

Difference versus Trend Stationary Models

We have encountered two ways to eliminate a trend. A trend stationary series can be transformed into a stationary series by removing the deterministic trend. A series with a unit root, sometimes called a **difference stationary** (DS) series, can be transformed into a stationary series by differencing. A serious problem is encountered when the

inappropriate method is used to eliminate trend. We saw an example of the problem in attempting to difference the equation $y_t = y_0 + a_1 t + \varepsilon_t$. Consider a more general trend stationary process of the form

$$A(L)y_t = a_0 + a_1 t + e_t$$

where the characteristic roots of the polynomial $A(L)$ are all outside the unit circle, and the expression e_t is allowed to have the form $e_t = B(L)\varepsilon_t$. Subtracting an estimate of the deterministic time trend yields a stationary and invertible ARMA model. However, if we use the notation of (4.11), the first difference of such a model yields

$$A(L)y_t^* = a_1 + (1 - L)B(L)\varepsilon_t$$

First differencing the *TS* process has introduced a noninvertible unit root process into the MA component of the model. Of course, the same problem is encountered in a model with a polynomial time trend.

In the same way, subtracting a deterministic time trend from a DS process is also inappropriate. For example, in the general trend plus irregular model of (4.9), subtracting $y_0 + a_0 t$ from each observation does not result in a stationary series since the stochastic portion of the trend is not eliminated.

Are There Business Cycles?

Traditional business cycle research decomposed real macroeconomic variables into a long-run (secular) trend and a cyclical component. The typical decomposition is illustrated by the hypothetical data in Figure 4.3. The secular trend, portrayed by the straight line, was deemed to be in the domain of growth theory. The slope of the trend line was thought to be determined by long-run factors such as technological growth, fertility, immigration, and educational attainment levels.

One source of the deviations from trend occurs because of the wavelike motion of real economic activity called the **business cycle**. Although the actual period of the cycle was never thought to be as regular as that depicted in the figure, the periods of prosperity and recovery were regarded to be as inevitable as the tides. The goal of monetary and fiscal policy was to reduce the amplitude of the cycle (measured by distance *ab*). In terms of our previous discussion, the trend is the nonstationary component, and the cyclical and irregular components are stationary.

Although there have been recessions and periods of high prosperity, the post-World War II experience taught us that business cycles do not have a regular period. Even so, there is a widespread belief that, over the long run, macroeconomic variables grow at a constant trend rate and that any deviations from trend are eventually eliminated by the *invisible hand*. The belief that trend is unchanging over time leads to the common practice of *detrending* macroeconomic data using a linear (or polynomial) deterministic regression equation. The lower portion of the figure shows the cycle and the noise (or irregular) component after detrending.

Nelson and Plosser (1982) challenged the traditional view by demonstrating that important macroeconomic variables tend to be DS rather than TS processes. They obtained time-series data for 13 important macroeconomic time series: real GNP, nominal GNP, industrial production, employment, unemployment rate, GNP deflator,

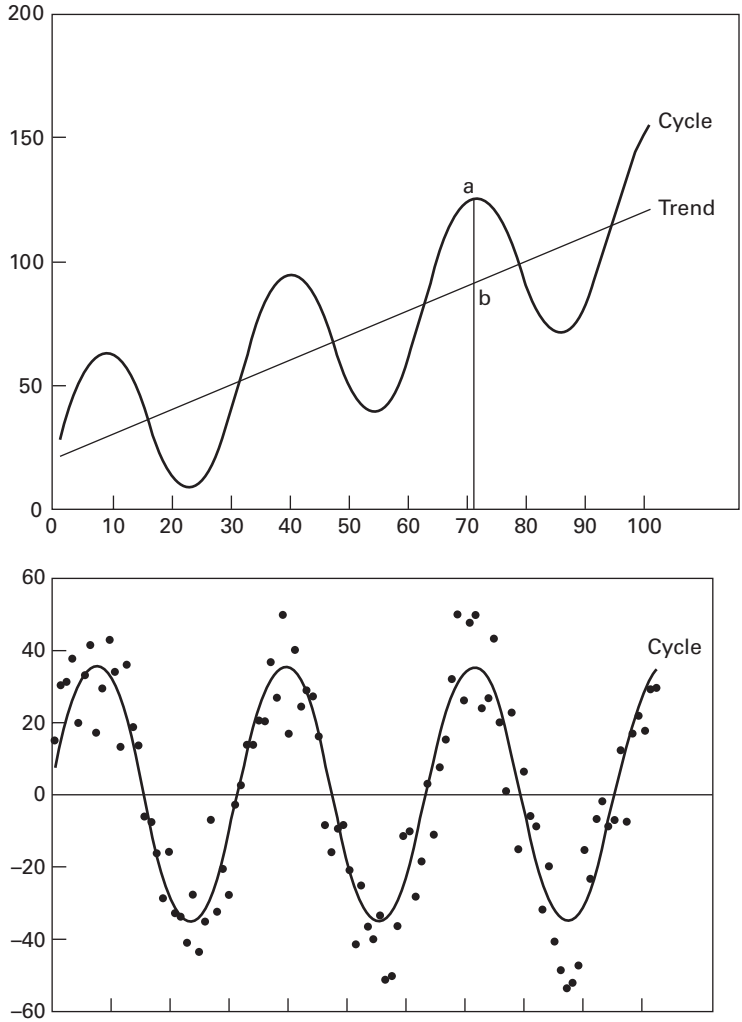


FIGURE 4.3 The Business Cycle?

consumer prices, wages, real wages, money stock, velocity, bond yields, and an index of common stock prices. The sample began as early as 1860 for consumer prices to as late as 1909 for GNP data and ended in 1970 for all of the series. Some of their findings are reported in Table 4.1. The first two columns report the first- and second-order autocorrelations of real and nominal GNPs, industrial production, and the unemployment rate. Notice that the autocorrelations of the first three of the series are strongly indicative of a unit root process. Although ρ_1 for the unemployment rate is 0.75, the second-order autocorrelation is less than 0.5.

First differences of the series yield the first- and second-order sample autocorrelations $r(1)$ and $r(2)$, respectively. Sample autocorrelations of the first differences are

Table 4.1 Selected Autocorrelations From Nelson and Plosser

	ρ_1	ρ_2	$r(1)$	$r(2)$	$d(1)$	$d(2)$
Real GNP	0.95	0.90	0.34	0.04	0.87	0.66
Nominal GNP	0.95	0.89	0.44	0.08	0.93	0.79
Industrial production	0.97	0.94	0.03	−0.11	0.84	0.67
Unemployment rate	0.75	0.47	0.09	−0.29	0.75	0.46

Notes:
[1]Full details of the correlogram can be obtained from Nelson and Plosser (1982), who report the first six sample autocorrelations.
[2]ρ_i, $r(i)$, and $d(i)$ refer to the ith-order autocorrelation coefficient for each series, for the first difference of the series, and for the detrended values of the series, respectively.

indicative of stationary processes. The evidence supports the claim that the data are generated from DS processes. Nelson and Plosser point out that the positive autocorrelation of differenced real and nominal GNP at lag 1 only is suggestive of an MA(1) process. To further strengthen the argument for DS processes, recall that differencing a TS process yields a noninvertible moving average process. None of the differenced series reported by Nelson and Plosser appears to have a unit root in the MA terms.

The results from fitting a linear trend to the data and forming sample autocorrelations of the residuals are given in the last two columns of the table. An interesting feature of the data is that the sample autocorrelations of the detrended data are reasonably high. This is consistent with the fact that detrending a DS series will not eliminate the nonstationarity. Notice that detrending the unemployment rate has *no effect* on the autocorrelations. The overall implication is that macroeconomic variables do not grow at a smooth long-run rate. Some macroeconomic shocks are of a permanent nature; the effects of such shocks are never eliminated.

The Trend in Real GDP

Another way to make the same point is to note that the real GDP series shown in Figure 4.1 has a clear trend. However, the tight fit of the estimated model might fool a researcher into thinking the series is actually stationary around the cubic trend line shown in Figure 4.1. Our eyes can be deceived because such trend lines are fit so as to make the observed residuals as small as possible. The ACF and PACF of the residuals from (4.1) are shown in Panel (a) of Figure 4.4. You can see that the ACF decays slowly while the PACF cuts to zero after one lag. In fact, this type of slow decay in the ACF is typical of a series with a stochastic trend. Thus, detrending the data does not seem to result in a stationary series. Panel (b) shows the ACF and PACF of the logarithmic change in real GDP. The ACF and PACF quickly converge to zero; after two lags, all autocorrelations and partial autocorrelations are not statistically different from zero. The estimated model for logarithmic change in real GDP ($\Delta lrgdp$) is

$$\begin{matrix} \Delta lrgdp_t = & 0.0049 & + & 0.3706 \Delta lrgdp_{t-1} \\ & (6.80) & & (6.44) \end{matrix}$$

Unlike the model of the deterministic trend, the residuals from this model all appear to be white noise. Thus, differencing is sufficient to remove the trend.

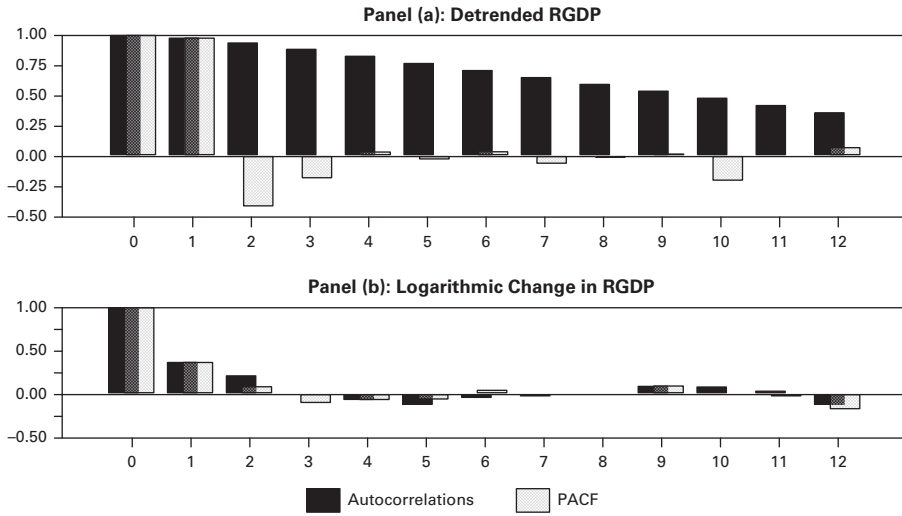


FIGURE 4.4 ACF and PACF

Rather than rely solely on an analysis of correlograms, it is possible to formally test whether a series is stationary. We examine such tests in the next several sections. The testing procedure is not as straightforward as it may seem. We cannot use the usual testing techniques because classical procedures all presume that the data are stationary. For now, it suffices to say that Nelson and Plosser are not able to reject the null hypothesis of a unit root. However, before we examine the tests for a unit root, it is important to note that the issue of nonstationarity also arises quite naturally in the context of the standard regression model.

3. UNIT ROOTS AND REGRESSION RESIDUALS

Consider the regression equation

$$y_t = a_0 + a_1 z_t + e_t \tag{4.12}$$

where the symbol e_t is used to indicate that the error term may be serially correlated.

The assumptions of the classical regression model necessitate that both the $\{y_t\}$ and $\{z_t\}$ sequences be stationary and that the errors have a zero mean and a finite variance. In the presence of nonstationary variables, there might be what Granger and Newbold (1974) call a **spurious regression**. A spurious regression has a high R^2- and t-statistics that appear to be significant, but the results are without any economic meaning. The regression output "looks good," but the least-squares estimates are not consistent and the customary tests of statistical inference do not hold. Granger and Newbold (1974) provide a detailed examination of the consequences of violating the stationarity assumption by generating two sequences, $\{y_t\}$ and $\{z_t\}$, as *independent* random walks using the formulas:

$$y_t = y_{t-1} + \varepsilon_{yt} \tag{4.13}$$

and

$$z_t = z_{t-1} + \varepsilon_{zt} \tag{4.14}$$

where ε_{yt} and ε_{zt} are white-noise processes that are independent of each other.

Granger and Newbold generated many such samples, and for each sample estimated, a regression in the form of (4.12). Since the $\{y_t\}$ and $\{z_t\}$ sequences are independent of each other, (4.12) is necessarily meaningless; any relationship between the two variables is spurious. Surprisingly, at the 5% significance level, they were able to reject the null hypothesis $a_1 = 0$ in approximately 75% of the cases. Of course, at the 5% level, a correctly sized test would yield rejections in only 5% of the regressions. Moreover, the regressions usually had very high R^2 values, and the estimated residuals exhibited a high degree of autocorrelation.

To explain the findings of Granger and Newbold, note that the regression equation (4.12) is necessarily meaningless if the residual series $\{e_t\}$ is nonstationary. Obviously, if the $\{e_t\}$ sequence has a stochastic trend, any error in period t never decays so that any deviation from the model is permanent. It is hard to imagine attaching any importance to an economic model having permanent errors. The simplest way to examine the properties of the $\{e_t\}$ sequence is to abstract from the intercept term a_0 and rewrite (4.12) as

$$e_t = y_t - a_1 z_t$$

If y_t and z_t are generated by (4.13) and (4.14), we can impose the initial conditions $y_0 = z_0 = 0$ so that

$$e_t = \sum_{i=1}^{t} \varepsilon_{yi} - a_1 \sum_{i=1}^{t} \varepsilon_{zi} \tag{4.15}$$

Clearly, the variance of the error becomes infinitely large as t increases. Moreover, the error has a permanent component in that $E_t e_{t+i} = e_t$ for all $i \geq 0$. Hence, the assumptions embedded in the usual hypothesis tests are violated so that any t-test, F-test, or R^2 values are unreliable. It is easy to see why the estimated residuals from a spurious regression will exhibit a high degree of autocorrelation. Updating (4.15), you should be able to demonstrate that the theoretical value of the correlation coefficient between e_t and e_{t+1} goes to unity as t increases.

Even though the true value of $a_1 = 0$, suppose that you estimate (4.12) and want to test the null hypothesis $a_1 = 0$. From (4.15), it should be clear that the error term is nonstationary. Yet, the assumption that the error term is a unit root process is inconsistent with the distributional theory underlying the use of OLS. This problem will not disappear in large samples. In fact, Phillips (1986) proves that the larger the sample, the more likely you are to falsely conclude that $a_1 \neq 0$.

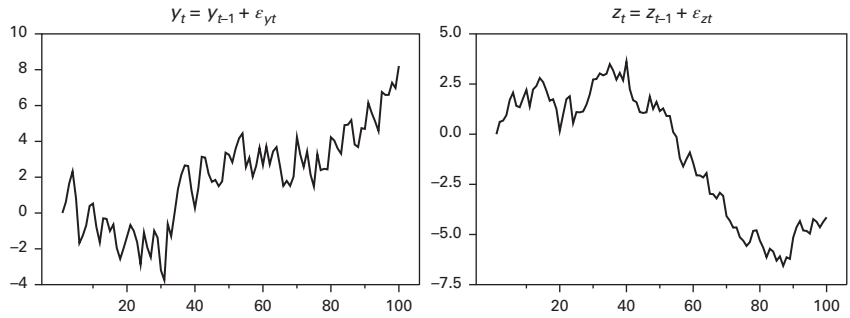
Worksheet 4.1 illustrates the problem of spurious regressions. The top two graphs show 100 realizations of the $\{y_t\}$ and $\{z_t\}$ sequences generated according to (4.13) and (4.14). Although $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are drawn from white-noise distributions, the realizations of the two sequences are such that y_{100} is positive and z_{100} is negative.

In the lower left panel, you can see that the regression of y_t on z_t captures the *within-sample* tendency of the sequences to move in opposite directions. The straight

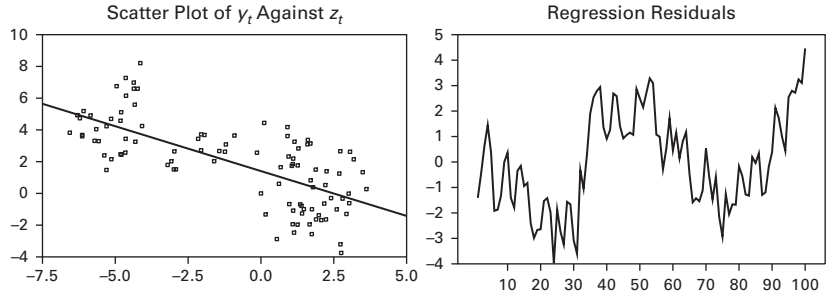
WORKSHEET 4.1

SPURIOUS REGRESSIONS: EXAMPLE 1

Consider the two random walk processes



Since both series are unit root processes with uncorrelated error terms, the regression of y_t on z_t is spurious. Given the realizations of $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$, it happens that y_t tends to increase as z_t tends to decrease. The regression line shown in the scatter plot of y_t on z_t captures this tendency. The correlation coefficient between y_t and z_t is -0.69 and a linear regression yields $y_t = 1.41 - 0.565z_t$. However, the residuals from the regression equation are nonstationary.

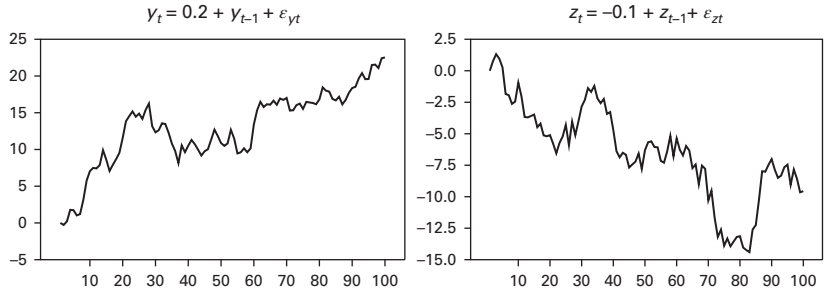


line shown in the scatter plot is the OLS regression line $y_t = 1.41 - 0.565z_t$. The correlation coefficient between $\{y_t\}$ and $\{z_t\}$ is -0.69. The residuals from this regression have a unit root; as such, the coefficients 1.41 and -0.565 are spurious. Worksheet 4.2 illustrates the same problem using two simulated random walk plus drift sequences: $y_t = 0.2 + y_{t-1} + \varepsilon_{yt}$ and $z_t = -0.1 + z_{t-1} + \varepsilon_{zt}$. The drift terms dominate so that for small values of t, it appears that $y_t = -2z_t$. As sample size increases, however, the cumulated sum of the errors (i.e., Σe_t) will pull the relationship further and further from -2.0. The scatter plot of the two sequences suggests that the R^2 statistic will be close to unity; in fact, R^2 is 0.93. However, as you can see in the last panel of

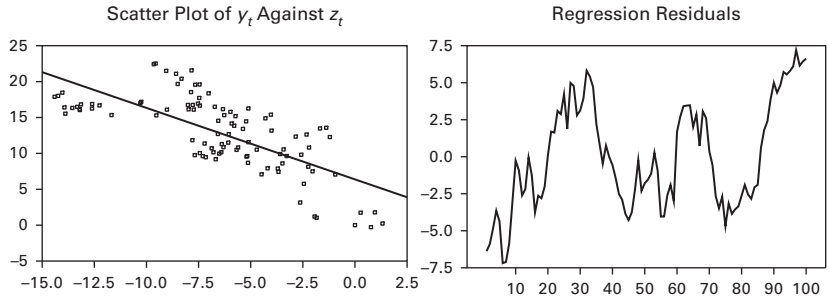
WORKSHEET 4.2

SPURIOUS REGRESSIONS: EXAMPLE 2

Consider the two random walk plus drift processes



Again, the $\{y_t\}$ and $\{z_t\}$ series are unit root processes with uncorrelated error terms so that the regression of y_t on z_t is spurious. Although it is the deterministic drift terms that cause the sustained increase in y_t and the overall decline in z_t, it appears that the two series are inversely related to each other. The residuals from the regression $y_t = 6.38 - 0.10z_t$ are nonstationary.



Worksheet 4.2, the residuals from the regression equation are nonstationary. All departures from this relationship are necessarily permanent.

The point is that the econometrician has to be very careful in working with nonstationary variables. In terms of (4.12), there are four cases to consider:

CASE 1

Both $\{y_t\}$ and $\{z_t\}$ are stationary. When both variables are stationary, the classical regression model is appropriate.

CASE 2

The $\{y_t\}$ and $\{z_t\}$ sequences are integrated of different orders. Regression equations using such variables are meaningless. For example, replace (4.14) with the stationary process $z_t = \rho z_{t-1} + \varepsilon_{zt}$ where $|\rho| < 1$. Now (4.15) is replaced by $e_t = \Sigma\varepsilon_{yi} - a_1\Sigma\rho^i\varepsilon_{zt-i}$. Although the expression $\Sigma\rho^i\varepsilon_{zt-i}$ is convergent, the $\{e_t\}$ sequence still contains a stochastic trend component.[2]

CASE 3

The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order, and the residual sequence contains a stochastic trend. This is the case in which the regression is spurious. The results from such spurious regressions are meaningless in that all errors are permanent. In this case, it is often recommended that the regression equation be estimated in first differences. Consider the first difference of (4.12):

$$\Delta y_t = a_1\Delta z_t + \Delta e_t$$

Since y_t, z_t, and e_t each contain unit roots, the first difference of each is stationary. Hence, the usual asymptotic results apply. Of course, if one of the trends is deterministic and the other is stochastic, first differencing each is not appropriate.

CASE 4

The nonstationary $\{y_t\}$ and $\{z_t\}$ sequences are integrated of the same order and the residual sequence is stationary. In this circumstance, $\{y_t\}$ and $\{z_t\}$ are **cointegrated**. A trivial example of a cointegrated system occurs if ε_{zt} and ε_{yt} are perfectly correlated. If $\varepsilon_{zt} = \varepsilon_{yt}$, then (4.15) can be set equal to zero (which is stationary) by setting $a_1 = 1$. To consider a more interesting example, suppose that both z_t and y_t are the random walk plus noise processes:

$$y_t = \mu_t + \varepsilon_{yt}$$
$$z_t = \mu_t + \varepsilon_{zt}$$

where ε_{yt} and ε_{zt} are white-noise processes and μ_t is the random walk process $\mu_t = \mu_{t-1} + \varepsilon_t$. Note that both $\{z_t\}$ and $\{y_t\}$ are $I(1)$ processes but that $y_t - z_t = \varepsilon_{yt} - \varepsilon_{zt}$ is stationary. The subtraction of z_t from y_t serves to nullify the stochastic trend.

All of Chapter 6 is devoted to the issue of cointegrated variables. For now, it is sufficient to note that pretesting the variables in a regression for nonstationarity is extremely important. Estimating a regression in the form of (4.12) is meaningless if cases 2 or 3 apply. If the variables are cointegrated, the results of Chapter 6 apply. The remainder of this chapter considers the formal test procedures for the presence of unit roots and/or deterministic time trends.

4. THE MONTE CARLO METHOD

As an applied researcher, you need to know whether a data series contains a trend and the best way to estimate the trend. You also need to avoid several critical mistakes. Clearly, you do not want to difference or detrend a stationary series. Moreover, you do not want to detrend a unit root process or difference a trend stationary process. Although the properties of a sample correlogram are useful tools for detecting the possible presence of unit roots or deterministic trends, the method is necessarily imprecise. What may appear as a unit root to one observer may appear as a stationary process to another. The problem is difficult because a near–unit root process will have the same shaped ACF as that of a process containing a trend. For example, the correlogram of a stationary AR(1) process such that $\rho_1 = 0.95$ will exhibit the type of gradual decay indicative of a nonstationary process. To illustrate some of the issues involved, suppose that we know a series is generated from the following first-order process:

$$y_t = a_1 y_{t-1} + \varepsilon_t \tag{4.16}$$

where $\{\varepsilon_t\}$ is white noise.

First, suppose that we wish to test the null hypothesis that $a_1 = 0$. Under the maintained null hypothesis of $a_1 = 0$, we can estimate (4.16) using OLS. The fact that ε_t is a white-noise process and that $|a_1| < 1$ guarantees that the $\{y_t\}$ sequence is stationary and that the estimate of a_1 is efficient. Calculating the standard error of the estimate of a_1, the researcher can use a t-test to determine whether a_1 is significantly different from zero.

The situation is quite different if we want to test the hypothesis $a_1 = 1$. Now, under the null hypothesis, the $\{y_t\}$ sequence is generated by the nonstationary process:

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i \tag{4.17}$$

Thus, if $a_1 = 1$, the variance becomes infinitely large as t increases. Under the null hypothesis, it is inappropriate to use classical statistical methods to estimate and perform significance tests on the coefficient a_1. If the $\{y_t\}$ sequence is generated as in (4.17), it is simple to show that the OLS estimate of (4.16) will yield a biased estimate of a_1. In Section 1, it was shown that the first-order autocorrelation coefficient in a random walk model is

$$\rho_1 = [(t-1)/t]^{0.5} < 1$$

Since the estimate of a_1 is directly related to the value of ρ_1, the estimated value of a_1 is biased to be below its true value of unity. The estimated model will mimic that of a stationary AR(1) process with a near unit root. Hence, the usual t-test cannot be used to test the hypothesis $a_1 = 1$.

Figure 4.5 shows the sample correlogram for a simulated random walk process. One hundred normally distributed random deviates were obtained so as to mimic the $\{\varepsilon_t\}$ sequence. Assuming $y_0 = 0$, the next 100 values in the $\{y_t\}$ sequence were calculated as $y_t = y_{t-1} + \varepsilon_t$. This particular correlogram is characteristic of most sample correlograms constructed from nonstationary data. The estimated value of ρ_1 is close

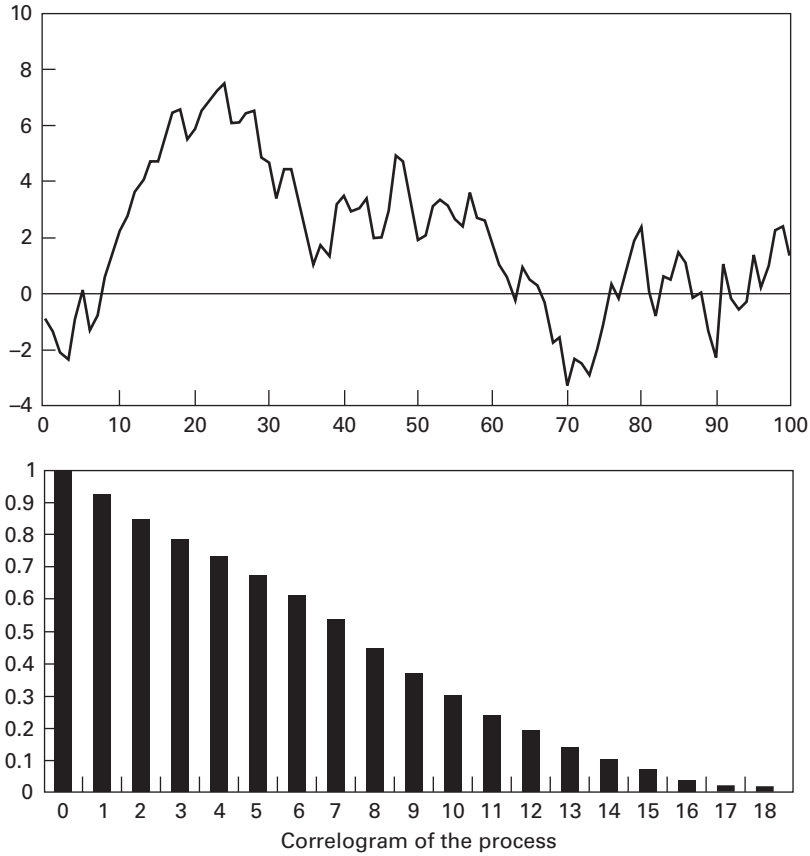


FIGURE 4.5 A Simulated Random Walk Process

to unity and the sample autocorrelations die out slowly. If we did not know the way in which the data were generated, inspection of Figure 4.5 might lead us to falsely conclude that the data were generated from a stationary process. With this particular data, estimates of an AR(1) model with and without an intercept yield (standard errors are in parentheses):

$$y_t = \underset{(0.030)}{0.9546} y_{t-1} + \varepsilon_t \qquad R^2 = 0.860 \tag{4.18}$$

$$y_t = 0.164 + \underset{(0.037)}{0.9247} y_{t-1} + \varepsilon_t \qquad R^2 = 0.864 \tag{4.19}$$

Examining (4.18), a careful researcher would not be willing to dismiss the possibility of a unit root since the estimated value of a_1 is only 1.5133 standard deviations from unity: $[(1 - 0.9546)/0.30 = 1.5133]$. We might correctly recognize that, under the null hypothesis of a unit root, the estimate of a_1 will be biased below unity. If we knew the true distribution of a_1 under the null of a unit root, we could perform such a significance test. Of course, if we did not know the true data-generating process, we might

estimate the model with an intercept. In (4.19), the estimate of a_1 is more than two standard deviations from unity: $(1 - 0.9247)/0.037 = 2.035$. However, it would be wrong to use this information to reject the null of a unit root. After all, the point of this section has been to indicate that such *t*-tests are inappropriate under the null of a unit root.

Fortunately, Dickey and Fuller (1979, 1981) devised a procedure to formally test for the presence of a unit root. Their methodology is similar to that used in constructing the data reported in Figure 4.5. Suppose that we generated thousands of random walk sequences and that, for each, we calculated the estimated value of a_1. Although most of the estimates would be close to unity, some would be further from unity than others. In performing this experiment, Dickey and Fuller found that in the presence of an intercept:

- 90% of the estimated values of a_1 are less than 2.58 standard errors from unity;
- 95% of the estimated values of a_1 are less than 2.89 standard errors from unity;
- 99% of the estimated values of a_1 are less than 3.51 standard errors from unity.

The application of these Dickey–Fuller *critical values* to tests for unit roots is straightforward. Suppose we did not know the true data-generating process and were trying to ascertain whether the data used in Figure 4.5 contained a unit root. Using these Dickey–Fuller statistics, we would not reject the null of a unit root in (4.19). The estimated value of a_1 is only 2.035 standard deviations from unity. In fact, if the true value of a_1 does equal unity, we should find the estimated value to be within 2.58 standard deviations from unity 90% of the time.

Be aware that stationarity necessitates $-1 < a_1 < 1$ or, equivalently, $a_1^2 < 1$. Thus, if the estimated value of a_1 is close to -1, you should also be concerned about nonstationarity. If we define $\gamma = a_1 - 1$, the equivalent restriction is $-2 < \gamma < 0$. In conducting a Dickey–Fuller test, *it is possible to check that the estimated value of* γ *is greater than* -2.[3] Nevertheless, with economic data, such a case is exceedingly rare. As such, almost all unit root tests are one-sided tests with the alternative hypothesis $\gamma < 0$.

Monte Carlo Experiments

The procedure that Dickey and Fuller (1979, 1981) used to obtain their critical values is typical of that found in the modern time-series literature. Hypothesis tests concerning the coefficients of nonstationary variables cannot be conducted using traditional *t*-tests or *F*-tests. The distributions of the appropriate test statistics are nonstandard and cannot be analytically evaluated. However, given the trivial cost of computer time, the nonstandard distributions can easily be derived using a Monte Carlo simulation.

A Monte Carlo experiment attempts to replicate an actual data-generating process (DGP) on a computer. To be more specific, you simulate a data set with the essential characteristics of the actual data in question. A Monte Carlo experiment generates a random sample of size T and the parameters and/or sample statistics of interest are calculated. This process is repeated N times (where N is a large number) so that the distribution of the desired parameters and/or sample statistics can be tabulated. These empirical distributions are used as estimates of the actual distributions.

All major statistical software packages have a built-in random number generator. The first step in a Monte Carlo experiment is to computer generate a set of random numbers (sometimes called pseudorandom numbers) from a given distribution. Of course, the numbers cannot be entirely random since all computer algorithms rely on a deterministic number-generating mechanism. However, the numbers are drawn so as to mimic a random process having some specified distribution. Usually, the numbers are designed to be normally distributed and serially uncorrelated. The idea is to use these numbers to represent one replication of the entire $\{\varepsilon_t\}$ sequence. If you want to know more about pseudorandom number generation, see Section 4.2 of the *Supplementary Manual*. The *Programming Manual* illustrates the Monte Carlo method for a number of different time-series models.

The second step is to construct the $\{y_t\}$ sequence using the random numbers and the parameters of the data-generating process. For example, Dickey and Fuller (1979, 1981) obtained 100 values for $\{\varepsilon_t\}$, set $a_1 = 1$, $y_0 = 0$ and calculated 100 values for $\{y_t\}$ according to (4.16). Once a series has been generated, the third step is to estimate the parameters of interest (such as the estimate of a_1 or the in-sample variance of the $\{y_t\}$ series).

The beauty of the method is that all important attributes of the constructed $\{y_t\}$ sequence are known to the researcher. For this reason, a Monte Carlo simulation is often referred to as an "experiment." The only problem is that the set of random numbers drawn is just one possible outcome. Obviously, the estimates in (4.18) and (4.19) are dependent on the values of the simulated $\{\varepsilon_t\}$ sequence. Different outcomes for $\{\varepsilon_t\}$ will yield different values of the simulated $\{y_t\}$ sequence.

This is why Monte Carlo studies perform many replications of the process outlined above. The fourth step is to replicate steps 1 and 3 thousands of times. The goal is to ensure that the statistical properties of the constructed $\{y_t\}$ sequence are in accordance with the true distribution. Thus, for each replication, the parameters of interest are tabulated and critical values (or confidence intervals) obtained. As such, the properties of your data can be compared to the properties of the simulated data so that hypothesis tests can be performed.

For our purposes, it suffices to say that the use of the Monte Carlo method is warranted by the Law of Large Numbers. Consider the simplest case where v_t is an identically and independently distributed (i.i.d.) random number with mean μ and variance σ^2 so that

$$v_t \sim (\mu, \sigma^2)$$

The sample mean constructed by using T observations of the $\{v_t\}$ sequence is

$$\overline{v} = (1/T)\sum_{t=1}^{T} v_t$$

By the Law of Large Numbers, as the sample size T grows sufficiently large, $\overline{v}$ converges to the true mean μ. Hence, the sample mean $\overline{v}$ is an unbiased estimate of the population mean. This is the justification for using the Dickey–Fuller critical values to test the hypothesis $a_1 = 1$. Moreover, if the draws are independent and the sample size

T grows sufficiently large, the distribution of $\bar{v}$ approaches a normal distribution with mean μ and variance σ^2/T.[4]

An important limitation of a Monte Carlo experiment is that the results are specific to the assumptions used to generate the simulated data. If you change the sample size, include (or delete) an additional parameter in the data-generating process, or use alternative initial conditions, an entirely new simulation needs to be performed. Moreover, the precision of your estimates depends on the number of replications you use. Oftentimes, you do not need many replications to obtain a good estimate of a population mean. However, it is necessary to use many thousands of replications to obtain good estimates of critical values. Nevertheless, you should be able to envision many applications of Monte Carlo experiments. As discussed in Hendry, Neale, and Ericsson (1990), they are particularly helpful for studying the small-sample properties of time-series data. As you will see shortly, Monte Carlo experiments are the workhorse of many tests used in modern time-series analysis.

Example of the Monte Carlo Method

Suppose you did not know the probability distribution for the sum of the roll of two dice. One way to calculate the probability distribution would be to buy a pair of dice and roll them several thousand times. If the dice were fair, you would find that a sum on your rolls would approximate this result:

Sum	2	3	4	5	6	7	8	9	10	11	12
Percentage	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Instead of actually rolling the dice, you can easily replicate the experiment on a computer. You could draw a random number from a uniform [0, 1] distribution to replicate the roll of the first die. If the computer-generated number falls within the interval [0, 1/6], set the variable $r_1 = 1$. Similarly, if the number falls within the interval [1/6, 2/6], set $r_1 = 2$, and so on. In this way, r_1 will be some integer 1 through 6, each with a probability 1/6. Next, draw a second number from the same uniform [0, 1] distribution to represent the roll of die 2 (r_2). You complete your first Monte Carlo replication by computing the sum $r_1 + r_2$. If you compute several thousand such sums, the sample distribution of the sums will approximate the true distribution.

Of course, more complicated experiments are possible. It is interesting to note that this method was used to reform a standard recommendation at the blackjack tables. At one time, the recommendation was to "stick" if the dealer shows a 2 or a 3 and you hold a 12. Monte Carlo experiments of a game of blackjack showed that this recommendation was incorrect. Now, a sharp blackjack player will take another card in these circumstances.

Generating the Dickey–Fuller Distribution

We need to modify the procedure above only slightly to obtain the Dickey and Fuller (1979) distribution. To generate the distribution for a sample size of 100, we can

perform the following steps:

STEP 1: First, we need a set of random numbers to represent the $\{\varepsilon_t\}$ sequence. If we use the usual set of assumptions, we can draw a set of 100 random numbers from a standard normal distribution. Of course, the Monte Carlo method would allow us to experiment with other distributions.

STEP 2: We need to generate the sequence $y_t = y_{t-1} + \varepsilon_t$. Note that we need to initialize the value of y_0. Once we draw the value of ε_1, we cannot construct y_1 without positing some value for y_0. However, we do not want the results to be sensitive to the initial value chosen for the series. Two slightly different procedures are used to purge the effects of the initial condition from the Monte Carlo results. First, you can initialize the value of y_0 to equal the unconditional mean of the $\{y_t\}$ sequence. Alternatively, suppose you want to generate T values of the $\{y_t\}$ sequence. You can pick an initial condition for y_0 and then generate the next $T + 50$ realizations. Discard the first 50 realizations and use only the last T values. The idea is that the effect of the initial condition will dissipate after 50 periods.

STEP 3: We need to estimate the model under the alternative hypothesis. As such, we estimate an equation of the form $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$. Obtain the t-statistic for the null hypothesis $\gamma = 0$. Note that the data are generated under the null hypothesis of a unit root and estimated under the alternative hypothesis.

STEP 4: Repeat steps 1–3, 10,000 or more times. If you use a sample size such that $T = 100$, you should obtain something very similar to the Dickey–Fuller τ_μ distribution plotted in Figure 4.6. Of course, you will not obtain the exact numbers used in the figure since you will be using a different set of random numbers.

The data used to draw Figure 4.6 contains 10,000 replications. Additional replications would reveal a somewhat smoother probability distribution. As you might expect,

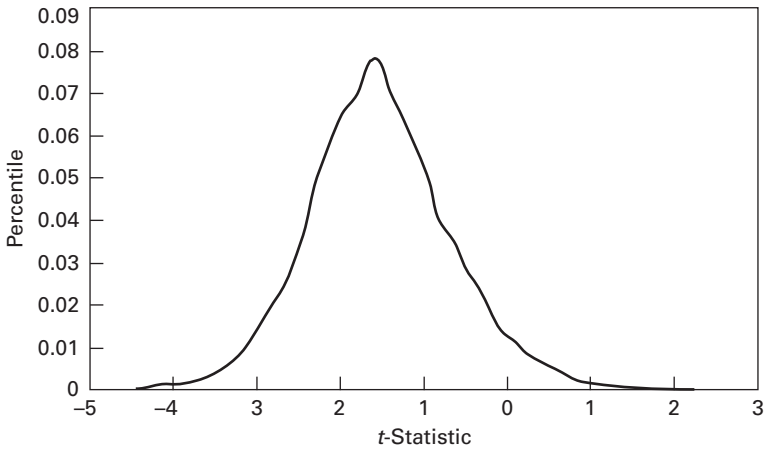


FIGURE 4.6 The Dickey–Fuller Distribution

the mean of the distribution is far below zero. The mean of the t-statistics shown in the figure is -1.53. The distribution of t-statistics for the null hypothesis $\gamma = 0$ is only slightly different from those reported by Dickey and Fuller; about 95% are more than -2.89 and 99% are more than -3.51. Hence, if you estimate a model in the form $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$ and find that the t-statistic for the null hypothesis $\gamma = 0$ is -3.00, you can reject the null hypothesis of a unit root at the 5%, but not at the 1%, level of significance. We will encounter a number of additional applications of Monte Carlo experiments throughout the text. Additional details of Monte Carlo and bootstrapping techniques are discussed in Section 4.3 of the *Supplementary Manual* and in the *Programming Manual*.

5. DICKEY–FULLER TESTS

The last section outlined a simple procedure to determine whether $a_1 = 1$ in the model $y_t = a_1 y_{t-1} + \varepsilon_t$. Begin by subtracting y_{t-1} from each side of the equation in order to write the equivalent form: $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$ where $\gamma = a_1 - 1$. Of course, testing the hypothesis $a_1 = 1$ is equivalent to testing the hypothesis $\gamma = 0$. Dickey and Fuller (1979) actually consider three different regression equations that can be used to test for the presence of a unit root:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \tag{4.20}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t \tag{4.21}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t \tag{4.22}$$

The difference between the three regressions concerns the presence of the deterministic elements a_0 and $a_2 t$. The first is a pure random walk model, the second adds an intercept or a *drift* term, and the third includes both a drift and a linear time trend.

The parameter of interest in all the regression equations is γ; if $\gamma = 0$, the $\{y_t\}$ sequence contains a unit root. The test involves estimating one (or more) of the equations above using OLS in order to obtain the estimated value of γ and the associated standard error. Comparing the resulting t-statistic with the appropriate value reported in the Dickey–Fuller tables allows the researcher to determine whether to accept or reject the null hypothesis $\gamma = 0$.

Recall that, in (4.18), the estimate of $y_t = a_1 y_{t-1} + \varepsilon_t$ was such that $a_1 = 0.9546$ with a standard error of 0.030. Clearly, the OLS regression in the form $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$ will yield an estimate of γ equal to -0.0454 with the same standard error of 0.030. Hence, the associated t-statistic for the hypothesis $\gamma = 0$ is -1.5133 (i.e., $-0.0454/0.03 = -1.5133$).

The methodology is precisely the same regardless of which of the three forms of the equations is estimated. However, be aware that the critical values of the t-statistics do depend on whether an intercept and/or time trend is included in the regression equation. In their Monte Carlo study, Dickey and Fuller (1979) found that the critical values for

$\gamma = 0$ depend on the form of the regression and sample size. The statistics called τ, τ_μ, and τ_τ are the appropriate statistics to use for (4.20–4.22), respectively.

Now, look at Table A in the *Supplementary Manual*. With 100 observations, there are three different critical values for the t-statistic $\gamma = 0$. For a regression without the intercept and trend terms ($a_0 = a_2 = 0$), use the section labeled τ. With 100 observations, the critical values for the t-statistic are -1.61, -1.95, and -2.60 at the 10%, 5%, and 1% significance levels, respectively. Thus, in the hypothetical example with $\gamma = -0.0454$ and a standard error of 0.03 (so that $t = -1.5133$), it is not possible to reject the null of a unit root at conventional significance levels. Note that the appropriate critical values depend on sample size. As in most hypothesis tests, for any given level of significance, the critical values of the t-statistic decrease as sample size increases.

Including an intercept term but not a trend term (only $a_2 = 0$) necessitates the use of the critical values in the section labeled τ_μ. Estimating (4.19) in the form $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$ necessarily yields a value of γ equal to $(0.9247 - 1) = -0.0753$ with a standard error of 0.037. The appropriate calculation for the t-statistic yields $-0.0753/0.037 = -2.035$. If we read from the appropriate row of Table A, with the same 100 observations, the critical values are -2.58, -2.89, and -3.51 at the 10%, 5%, and 1% significance levels, respectively. Again, the null of a unit root cannot be rejected at conventional significance levels. Finally, with both intercept and trend, use the critical values in the section labeled τ_τ; now, the critical values are -3.45 and -4.04 at the 5% and 1% significance levels, respectively. The equation was not estimated using a time trend; inspection of Figure 4.5 indicates that there is little reason to include a deterministic trend in the estimating equation.

As discussed in Section 7, these critical values are unchanged if (4.20–4.22) are replaced by the autoregressive processes:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \tag{4.23}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \tag{4.24}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \tag{4.25}$$

Tests including lagged changes are called **augmented** Dickey–Fuller tests and the same τ, τ_μ, and τ_τ statistics are all used to test the hypotheses $\gamma = 0$. Dickey and Fuller (1981) provide three additional F-statistics (called ϕ_1, ϕ_2, and ϕ_3) to test joint hypotheses on the coefficients. Using (4.21) or (4.24), the null hypothesis $\gamma = a_0 = 0$ is tested using the ϕ_1 statistic. Including a time trend in the regression—so that (4.22) or (4.25) is estimated—the joint hypothesis $a_0 = \gamma = a_2 = 0$ is tested using the ϕ_2 statistic and the joint hypothesis $\gamma = a_2 = 0$ is tested using the ϕ_3 statistic.

Table 4.2 Summary of the Dickey–Fuller Tests

Model	Hypothesis	Test Statistic	Critical Values for 95% and 99% Confidence Intervals
$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ	−3.45 and −4.04
	$\gamma = a_2 = 0$	ϕ_3	6.49 and 8.73
	$a_0 = \gamma = a_2 = 0$	ϕ_2	4.88 and 6.50
$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ_μ	−2.89 and −3.51
	$a_0 = \gamma = 0$	ϕ_1	4.71 and 6.70
$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$	$\gamma = 0$	τ	−1.95 and −2.60

Note: Critical values are for a sample size of 100.

The ϕ_1, ϕ_2, and ϕ_3 statistics are constructed in exactly the same way as ordinary F-tests:

$$\phi_i = \frac{[\text{SSR(restricted)} - \text{SSR(unrestricted)}]/r}{\text{SSR(unrestricted)}/(T-k)}$$

where SSR(restricted) and SSR(unrestricted) = the sums of the squared residuals from the restricted and unrestricted models, respectively,
r = number of restrictions,
T = number of usable observations, and
k = number of parameters estimated in the unrestricted model.

Hence, $T - k$ = degrees of freedom in the unrestricted model.

Comparing the calculated value of ϕ_i to the appropriate value reported in Dickey and Fuller (1981) allows you to determine the significance level at which the restriction is binding. The null hypothesis is that the data are generated by the restricted model, and the alternative hypothesis is that the data are generated by the unrestricted model. If the restriction is not binding SSR(restricted) should be close to SSR(unrestricted) and ϕ_i should be small; hence, large values of ϕ_i suggest a binding restriction and a rejection of the null hypothesis. Thus, if the calculated value of ϕ_i is smaller than that reported by Dickey and Fuller, you can accept the restricted model (i.e., you do not reject the null hypothesis that the restriction is not binding). If the calculated value of ϕ_i is larger than that reported by Dickey and Fuller, you can reject the null hypothesis and conclude that the restriction is binding. The critical values of the three ϕ_i statistics are reported in Table B in the *Supplementary Manual*. The complete set of test statistics and their critical values for a sample size of 100 is summarized in Table 4.2.

An Example

To illustrate the use of the various test statistics, Dickey and Fuller (1981) use quarterly values of the logarithm of the Federal Reserve Board's Production Index over the

1950Q1–1977Q4 period to estimate the following three equations:

$$\Delta y_t = 0.52 + 0.00120t - 0.119y_{t-1} + 0.498\Delta y_{t-1} + \varepsilon_t \qquad \text{SSR} = 0.056448$$
$$(0.15)\ (0.00034) \quad (0.033) \qquad (0.081) \tag{4.26}$$

$$\Delta y_t = 0.0054 + 0.447\Delta y_{t-1} + \varepsilon_t \qquad \text{SSR} = 0.063211$$
$$(0.0025)\ (0.083) \tag{4.27}$$

$$\Delta y_t = 0.511\Delta y_{t-1} + \varepsilon_t \qquad \text{SSR} = 0.065966$$
$$(0.079) \tag{4.28}$$

where SSR = sum of squared residuals and standard errors are in parentheses.

To test the null hypothesis that the data are generated by (4.28) against the alternative that (4.26) is the "true" model, use the ϕ_2 statistic. Dickey and Fuller test the null hypothesis $a_0 = a_2 = \gamma = 0$ as follows. Note that the residual sums of squares of the restricted and unrestricted models are 0.065966 and 0.056448, respectively, and that the null hypothesis entails three restrictions. With 110 usable observations and 4 estimated parameters, the unrestricted model contains 106 degrees of freedom. Since $0.056448/106 = 0.000533$, the ϕ_2 statistic is given by

$$\phi_2 = (0.065966 - 0.056448)/[3(0.000533)] = 5.95$$

With 110 observations, the critical value of ϕ_2 calculated by Dickey and Fuller is 5.59 at the 2.5% significance level. Hence, it is possible to reject the null hypothesis of a random walk against the alternative that the data contain an intercept and/or a unit root and/or a deterministic time trend (i.e., rejecting $a_0 = a_2 = \gamma = 0$ means that one or more of these parameters does not equal zero).

Dickey and Fuller also test the null hypothesis $a_2 = \gamma = 0$ given the alternative of (4.26). If we now view (4.27) as the restricted model and (4.26) as the unrestricted model, the ϕ_3 statistic is calculated as

$$\phi_3 = (0.063211 - 0.056448)/[2(0.000533)] = 6.34$$

With 110 observations, Table B indicates that the critical value of ϕ_3 is 6.49 at the 5% significance level and 5.47 at the 10% significance level. At the 10% level, they reject the null hypothesis and accept the alternative that the series is TS. However, at the 5% level, the calculated value of ϕ_3 is smaller than the critical value of 6.49; at this significance level, they do not reject the null hypothesis. Hence, at the 5% significance level, they maintain the hypothesis that the series contains a unit root and/or a deterministic time trend.

To compare with the τ_τ test (i.e., the hypothesis that only $\gamma = 0$), note that

$$\tau_\tau = -0.119/0.033 = -3.61$$

so that it is possible to reject the null of a unit root at the 5% level.

A number of examples and tips about the test are given in Chapter 6 of the *Programming Manual* that accompanies the text.

6. EXAMPLES OF THE DICKEY–FULLER TEST

Section 2 reviewed the evidence reported by Nelson and Plosser (1982) suggesting that macroeconomic variables are DS rather than trend stationary. We are now in a position to consider their formal tests of the hypothesis. For each series under study, Nelson and Plosser estimated the regression in the form of (4.25):

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

The chosen lag lengths are reported in the column labeled p in Table 4.3. The estimated values a_0, a_2, and γ are reported in columns 3, 4, and 5, respectively.

Recall that the old school view of business cycles maintains that GNP and production levels are trend stationary rather than DS. An adherent to this view must assert that γ is different from zero; if $\gamma = 0$, the series has a unit root and is DS. Given the sample sizes used by Nelson and Plosser (1982), at the 0.05 level, the critical value of the t-statistic for the null hypothesis $\gamma = 0$ is -3.45. Thus, only if the estimated value of γ is more than 3.45 standard deviations from zero it is possible to reject the hypothesis that $\gamma = 0$. As can be seen from inspection of Table 4.3, the estimated values of γ for real GNP, nominal GNP, and industrial production are not statistically different from zero. Only the unemployment rate has an estimated value of γ that is significantly different from zero at the 0.05 level.

Quarterly Real U.S. GDP

Now use the data on the file RGDP.XLS to estimate the logarithmic change in real GDP as

$$\underset{}{\Delta lrgdp_t} = \underset{(1.58)}{0.1248} + \underset{(1.31)}{0.0001t} - \underset{(-1.49)}{0.0156 lrgdp_{t-1}} + \underset{(6.26)}{0.3663 \Delta lrgdp_{t-1}} \qquad (4.29)$$

Table 4.3 The Tests by Nelson and Plosser for Unit Roots

	p	a_0	a_2	γ	$\gamma + 1$
Real GNP	2	0.819	0.006	−0.175	0.825
		(3.03)	(3.03)	(−2.99)	
Nominal GNP	2	1.06	0.006	−0.101	0.899
		(2.37)	(2.34)	(−2.32)	
Industrial production	6	0.103	0.007	−0.165	0.835
		(4.32)	(2.44)	(−2.53)	
Unemployment rate	4	0.513	−0.000	−0.294*	0.706
		(2.81)	(−0.23)	(−3.55)	

Notes:
[1] p is the chosen lag length. Coefficients divided by their standard errors are in parentheses. Thus, entries in parentheses represent the t-test for the null hypothesis that a coefficient is equal to zero. Under the null of nonstationarity, it is necessary to use the Dickey–Fuller critical values. At the 0.05 significance level, the critical value for the t-statistic is −3.45.
[2] An (*) denotes significance at the 0.05 level. For real and nominal GNPs and industrial production, it is not possible to reject the null hypothesis $\gamma = 0$ at the 0.05 level. Hence, the unemployment rate appears to be stationary.
[3] The expression $\gamma + 1$ is the estimate of a_1.

The t-statistic on the coefficient for $lrgdp_{t-1}$ is -1.49. Table A indicates that, with 244 usable observations, the 10% and 5% critical values of τ_τ are about -3.13 and -3.43, respectively. As such, we cannot reject the null hypothesis of a unit root. The sample value of ϕ_3 for the null hypothesis $a_2 = \gamma = 0$ is 2.97. As Table B indicates that the 10% critical value is 5.39, we cannot reject the joint hypothesis of a unit root and no deterministic time trend. Since the sample value of ϕ_2 (equal to 17.61) far exceeds the 5% critical value of 4.75, we do not want to exclude the drift term. We can conclude that the growth rate of the real GDP series acts as a random walk plus drift plus the irregular term $0.3663\Delta lrgdp_{t-1}$. Additional details are contained in Section 4.4 of the *Supplementary Manual*.

Unit Roots and Purchasing Power Parity

Purchasing power parity (PPP) is a simple relationship linking national price levels and exchange rates. In its simplest form, PPP asserts that the rate of currency depreciation is approximately equal to the difference between domestic and foreign inflation rates. If p_t and p_t^* denote the logarithms of U.S. and foreign price levels and e_t denotes the logarithm of the dollar price of foreign exchange, PPP implies

$$e_t = p_t - p_t^* + d_t$$

where d_t represents the deviation from PPP in period t.

In applied work, p_t and p_t^* usually refer to national price indices in t relative to a base year, so that e_t refers to an index of the domestic currency price of foreign exchange relative to a base year. For example, if the U.S. inflation rate is 10% while the foreign inflation rate is 15%, the dollar price of foreign exchange should fall by approximately 5%. The presence of the term d_t allows for short-run deviations from PPP.

Because of its simplicity and intuitive appeal, PPP has been used extensively in theoretical models of exchange rate determination. However, as in the well-known Dornbusch (1976) "overshooting" model, real economic shocks, such as productivity or demand shocks, can cause permanent deviations from PPP. For our purposes, the theory of PPP serves as an excellent vehicle to illustrate many time-series testing procedures. One test of long-run PPP is to determine whether d_t is stationary. After all, if the deviations from PPP are nonstationary (i.e., if the deviations are permanent in nature), we can reject the theory. Note that PPP does allow for persistent deviations; the autocorrelations of the $\{d_t\}$ sequence need not be zero. One popular testing procedure is to define the "real" exchange rate in period t:

$$r_t \equiv e_t + p_t^* - p_t$$

Long-run PPP is said to hold if the $\{r_t\}$ sequence is stationary. For example, in Enders (1988), I constructed real exchange rates for three major U.S. trading partners: Germany, Canada, and Japan. The data were divided into two periods: January 1960 to April 1971 (representing the fixed exchange rate period) and January 1973 to November 1986 (representing the flexible exchange rate period). Each nation's Wholesale Price Index (WPI) was multiplied by an index of the U.S. dollar price of the foreign currency and then divided by the U.S. WPI. The log of the constructed series is the $\{r_t\}$ sequence.

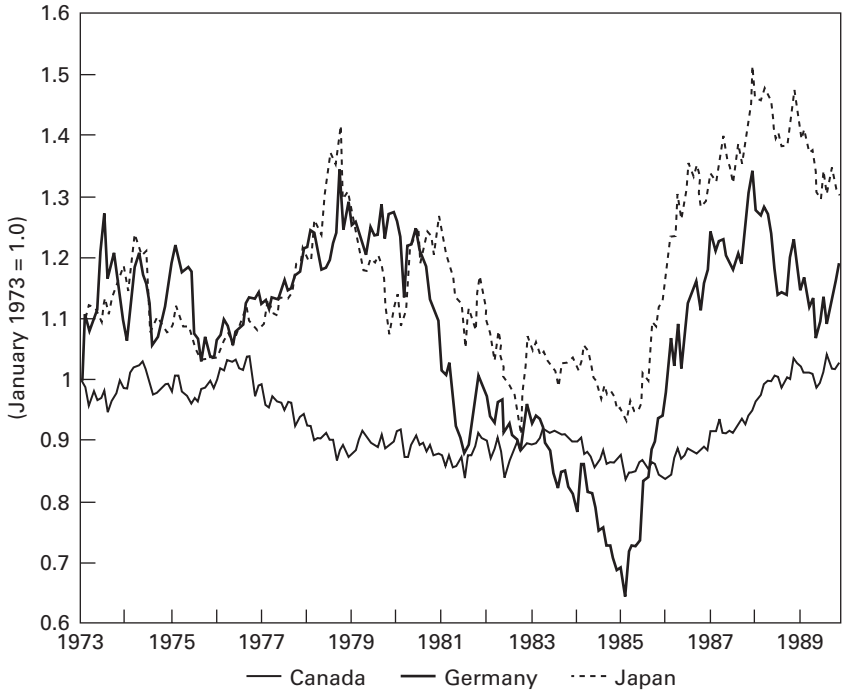


FIGURE 4.7 Real Exchange Rates

A critical first step in any econometric analysis is to visually inspect the data. The plots of the three real exchange rate series during the flexible exchange rate period (through 1989) are shown in Figure 4.7. Each series seems to meander in a fashion characteristic of a random walk process. Notice that there is little visual evidence of explosive behavior or a deterministic time trend. The autocorrelation function for all of the series in the analysis look similar to that in Figure 4.5. In particular, the autocorrelation functions show little tendency to decay while the autocorrelations of the first differences display the classic pattern of a stationary series.

To formally test for the presence of a unit root in the real exchange rates, augmented Dickey–Fuller tests of the form given by (4.24) were conducted. The regression $\Delta r_t = a_0 + \gamma r_{t-1} + \beta_2 \Delta r_{t-1} + \beta_3 \Delta r_{t-2} + \cdots + \varepsilon_t$ was estimated based on the following considerations:

1. The theory of PPP does not allow for a deterministic time trend. Any such findings would refute the theory as posited. Given that the series all decline throughout the early 1980s and all increase during the middle to late 1980s, there is no reason to entertain the notion of trend stationarity. As such, the expression $a_2 t$ was not included in the estimating equation.
2. For the fixed exchange rate period, various lag length tests indicated that all values of β_i could be set equal to zero for all three countries. However, different lag length tests yielded ambiguous results for the flexible exchange rate

period. Lag length tests indicated that β_{11} was statistically different from zero for all three countries. In contrast, *F*-tests and the SBC selected two lags for Germany and Japan and no lagged changes for Canada. As such, for the flexible rate period, the Dickey–Fuller tests were conducted using two different lag lengths for each country.

For the Canadian case during the 1973–1986 period, the *t*-statistic for the null hypothesis that $\gamma = 0$ is -1.42 using no lags and -1.51 using all 11 lags. Given the critical value of the τ_μ statistic, it is not possible to reject the null of a unit root in the Canadian/U.S. real exchange rate series. Hence, PPP fails for the Canadian–U.S. case. In the 1960–1971 period, the calculated value of the *t*-statistic is -1.59; again, it is possible to conclude that PPP fails.

Table 4.4 reports the results of all six estimations using the short lag lengths suggested by the *F*-tests and the SBC. Notice the following properties of the estimated models:

1. For all six models, it is not possible to reject the null hypothesis that PPP fails. As can be seen from the third column of Table 4.4, the absolute value of the *t*-statistic for the null $\gamma = 0$ is never more than 1.59. The economic interpretation is that real productivity and/or demand shocks have had a permanent influence on real exchange rates.
2. As measured by the sample SD, real exchange rates were far more volatile in the 1973–1986 period than in the 1960–1971 period. Moreover, as measured by the standard error of the estimate (SEE), real exchange rate volatility is

Table 4.4 Real Exchange Rate Estimation

	γ[1]	H_0: $\gamma = 0$[2]	Lags	Mean[3]	ρ/DW	*F*	SD/SEE
1973–1986							
Canada	−0.022	$t = -1.42$	0	1.05	0.059	0.194	5.47
	(0.016)				1.88		1.16
Japan	−0.047	$t = -0.64$	2	1.01	−0.007	0.226	10.44
	(0.074)				2.01		2.81
Germany	−0.027	$t = -0.28$	2	1.11	−0.014	0.858	20.68
	(0.076)				2.04		3.71
1960–1971							
Canada	−0.031	$t = -1.59$	0	1.02	−0.107	0.434	0.014
	(0.019)				2.21		0.004
Japan	−0.030	$t = -1.04$	0	0.98	0.046	0.330	0.017
	(0.028)				1.98		0.005
Germany	−0.016	$t = -1.23$	0	1.01	0.038	0.097	0.026
	(0.012)				1.93		0.004

Notes:
[1]Standard errors are in parentheses.
[2]Entries are the *t*-statistic for the hypothesis $\gamma = 0$.
[3]Mean is the sample mean of the series. SD is the standard deviation of the real exchange rate. SEE is the estimated standard deviation of the residuals (i.e., the standard error of the estimate). *F* is the significance level of the test that lags 2 (or 3) through 12 can be excluded. DW is the Durbin–Watson statistic for first-order serial correlation, and ρ is the estimated autocorrelation coefficient.

associated with unpredictability. The SEE during the flexible exchange rate period is several hundred times that of the fixed rate period. It seems reasonable to conclude that the change in the exchange rate regime (i.e., the end of Bretton Woods) affected the volatility of the real exchange rate.

3. Care must be taken to keep the appropriate null hypothesis in mind. Under the null of a unit root, classical test procedures are inappropriate, and we resort to the statistics tabulated by Dickey and Fuller. However, classical test procedures (which assume stationary variables) are appropriate under the null that the real rates are stationary. Thus, the following possibility arises: Suppose that the t-statistic in the Canadian case happened to be -2.16 instead of -1.42. If you used the Dickey–Fuller critical values, you would not reject the null of a unit root. Hence, you could conclude that PPP fails. However, under the null of stationarity (where we can use classical procedures), γ is more than two standard deviations from zero and you would *not reject the null of stationarity*.

This apparent dilemma commonly occurs when analyzing series with roots close to unity in absolute value. Unit root tests do not have much power in discriminating between characteristic roots close to unity and actual unit roots. The dilemma is only apparent since the two null hypotheses are quite different. It is perfectly consistent to maintain a null that PPP holds and not be able to reject a null that PPP fails! Notice that this dilemma does not arise for any of the series reported in Table 4.4; for each, it is not possible to reject a null of $\gamma = 0$ at conventional significance levels.

One way to circumvent this problem is to directly test the null hypothesis of stationarity against the alternative of nonstationarity. Kwiatowski, Phillips, Schmidt, and Shin (1992) show how to perform this type of test.

4. Looking at some of the diagnostic statistics, the F-statistics all indicate that it is appropriate to exclude lags 2 (or 3) through 12 from the regression equation. To reinforce the use of short lags, notice that the first-order correlation coefficient of the residuals (ρ) is low and that the Durbin–Watson statistic is close to two. It is interesting that the point estimates of the characteristic roots all indicate that real exchange rates are convergent. To obtain the characteristic roots, rewrite the estimated equations in the autoregressive form $r_t = a_0 + a_1 r_{t-1}$ or $r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-2}$. For the four AR(1) models, the point estimates of the slope coefficients are all less than unity. In the post-Bretton Woods period (1973–1986), the point estimates of the characteristic roots of Japan's second-order process are 0.931 and 0.319; for Germany, the roots are 0.964 and 0.256. Yet, this is precisely what we would expect if PPP fails; under the null of a unit root, we know that γ is biased downward.

To update the study, the file PANEL.XLS contains quarterly values of the real effective exchange rates (CPI based) for Australia, Canada, France, Germany, Japan, Netherlands, the United Kingdom, and the United States over the 1980Q1–2013Q1 period. These are multilateral (not bilateral) real exchange rates. As an exercise, you

should use these data to verify that very little has changed. You should find that only, for France and the Netherlands, is it possible to reject a unit root in the real exchange rate at the 5% significance level. Try not to peek; for each country, the estimated value of γ and the appropriate lag length are reported in Table 4.8.

7. EXTENSIONS OF THE DICKEY–FULLER TEST

Not all time-series variables can be well represented by the first-order autoregressive process $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$. It is possible to use the Dickey–Fuller tests in higher-order equations such as (4.23–4.25). Consider the pth order autoregressive process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t$$

To best understand the methodology of the **augmented Dickey–Fuller** (ADF) test, add and subtract $a_p y_{t-p+1}$ to obtain

$$\begin{aligned} y_t = {} & a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} \\ & + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + \varepsilon_t \end{aligned}$$

Next, add and subtract $(a_{p-1} + a_p) y_{t-p+2}$ to obtain

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots - (a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + \varepsilon_t$$

Continuing in this fashion, we obtain

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \tag{4.30}$$

where $\gamma = -\left(1 - \sum_{i=1}^{p} a_i\right)$ and $\beta_i = \sum_{j=i}^{p} a_j$

In (4.30), the coefficient of interest is γ; if $\gamma = 0$, the equation is entirely in first differences and, so, has a unit root. We can test for the presence of a unit root using the same Dickey–Fuller statistics discussed earlier. Again, the appropriate statistic to use depends on the deterministic components included in the regression equation. Without an intercept or a trend, use the τ statistic; with only the intercept, use the τ_μ statistic; and with both intercept and trend, use the τ_τ statistic. It is worthwhile pointing out that the results here are perfectly consistent with our study of difference equations in Chapter 1. If the coefficients of a difference equation sum to 1, *at least* one characteristic root is unity. Here, if $\Sigma a_i = 1$, $\gamma = 0$, and the system has a unit root.

Note that the Dickey–Fuller tests assume that the errors are independent and have a constant variance. This raises six important problems related to the fact that we do not know the true data-generating process:

1. We cannot properly estimate γ and its standard error unless all of the autoregressive terms are included in the estimating equation. Clearly, the simple regression $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$ is inadequate to this task if (4.30) is the true

data-generating process. Since the true order of the autoregressive process is unknown, the problem is to select the appropriate lag length.

2. The DGP may contain both autoregressive and moving average components. We need to know how to conduct the test if the order of the moving average terms (if any) is unknown.
3. The Dickey–Fuller test considers only a single unit root. However, a pth order autoregression has p characteristic roots; if there are $d \leq p$ unit roots, the series needs to be differenced d times to achieve stationarity.
4. As we saw in Chapter 2, there may be roots that require first differences and others that necessitate seasonal differencing. We need to develop a method that can distinguish between these two types of unit root processes.
5. There might be structural breaks in the data. As shown in Section 8, such breaks can impart an apparent trend to the data.
6. It might not be known whether an intercept and/or time trend belongs in (4.30). Section 9 is concerned with the issue of the appropriate deterministic regressors. (Additional details are given in Section 4.4 entitled "Determinants of the Deterministic Regressors" in the *Supplementary Manual*.)

Selection of the Lag Length

It is important to use the correct number of lags in conducting a Dickey–Fuller test. Too few lags mean that the regression residuals do not behave like white-noise processes. The model will not appropriately capture the actual error process so that γ and its standard error will not be well estimated. Including too many lags reduces the power of the test to reject the null of a unit root since the increased number of lags necessitates the estimation of additional parameters and a loss of degrees of freedom. The degrees of freedom decrease since the number of parameters estimated has increased and the number of usable observations has decreased. (We lose one observation for each additional lag included in the autoregression.) As such, the presence of unnecessary lags will reduce the power of the Dickey–Fuller test to detect a unit root. In fact, an augmented Dickey–Fuller test may indicate a unit root for some lag lengths but not for others.

How does a careful researcher select the appropriate lag length in such circumstances? One approach is the **general-to-specific** methodology. The idea is to start with a relatively long lag length and pare down the model by the usual t-test and/or F-tests. For example, one could estimate equation (4.30) using a lag length of p^*. If the t-statistic on lag p^* is insignificant at some specified critical value, reestimate the regression using a lag length of $p^* - 1$. Repeat the process until the last lag is significantly different from zero. In the pure autoregressive case, such a procedure will yield the true lag length with an asymptotic probability of unity, provided the initial choice of lag length includes the true length. Using seasonal data, the process is a bit different. For example, using quarterly data, one could start with 3 years of lags ($p = 12$). If the t-statistic on lag 12 is insignificant at some specified critical value and if an F-test indicates that lags 9–12 are also insignificant, move to lags 1–8. Repeat the process for lag 8 and lags 5–8 until a reasonable lag length has been determined.

Once a tentative lag length has been determined, diagnostic checking should be conducted. As always, plotting the residuals is a most important diagnostic tool. There should not appear to be any strong evidence of structural change or serial correlation. Moreover, the correlogram of the residuals should appear to be white noise. The Ljung–Box Q-statistic should not reveal any significant autocorrelations among the residuals. It is inadvisable to use the alternative procedure of beginning with the most parsimonious model and continuing to add lags until the first insignificant lag is found. Monte Carlo studies show that this procedure is biased toward selecting a value of p that is less than the true value.

As long as the regression equation does not omit a deterministic regressor present in the data-generating process, it is possible to perform lag length tests using t-tests or F-tests. The rationale follows from an important result proved by Sims, Stock, and Watson (1990). We will have cause to refer to several of the results of their paper. Here is the key finding of interest:

> **Rule 1:** Consider a regression equation containing a mixture of $I(1)$ and $I(0)$ variables such that the residuals are white noise. If the model is such that the coefficient of interest can be written as a coefficient on zero-mean stationary variables, then asymptotically, the OLS estimator converges to a normal distribution. As such, a t-test is appropriate.

Although this rule refers to any regression equation estimated by OLS, it applies directly to unit root tests. As shown above, the pth-order autoregressive process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t$$

can be written as

$$\Delta y_t = a_0 \gamma y_{t-1} + \beta_2 \Delta y_{t-1} + \beta_3 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p+1} + \varepsilon_t \tag{4.31}$$

From Rule 1, all the coefficients on the expressions Δy_{t-i} converge to t-distributions. As such, groups of these coefficients will converge to an F-distribution. Hence, you can perform a test of the form $\beta_i = \beta_{i+1} = \cdots = \beta_p = 0$ using an F-test. Nevertheless, under the null hypothesis of a unit root, the value of γ multiplies a nonstationary variable. As such, a test of $\gamma = 0$ *cannot* be conducted using a standard t-test.

In addition to the use of F-tests and t-tests, it is also possible to determine the lag length using an information criterion such as the AIC or SBC. Of course, in very large samples with normally distributed errors, the methods should all select the same lag length. In practice, the SBC will select a more parsimonious model than will either the AIC or t-tests. Nevertheless, whichever method is used, the researcher must ensure that residuals act as white-noise processes.

An Example: In order to illustrate the various procedures to select the lag length for an augmented Dickey–Fuller test, 200 realizations of the following unit root process were generated

$$\Delta y_t = 0.5 + 0.5 \Delta y_{t-1} + 0.2 \Delta y_{t-3} + \varepsilon_t$$

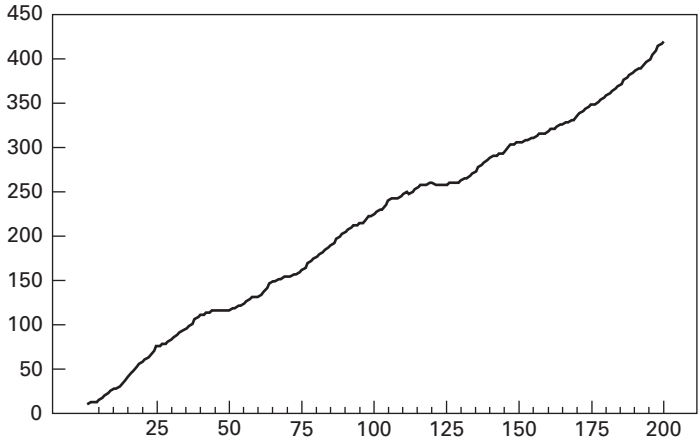


FIGURE 4.8 Unit Root Plus Drift

Notice that the $\{y_t\}$ sequence contains a single unit root and that the appropriate lag length is 3. The drift parameter gives the series the decidedly increasing pattern shown in Figure 4.8. (You can follow along using the data on the file LAGLENGTH.XLS.) Pretend that you do not know the actual DGP. As such, the time path of the sequence allows for two possible DGPs; the series may be trend stationary or a unit root process containing a drift term. Hence, the null hypothesis is that of a unit root process containing a drift against the alternative of a trend stationary process. The appropriate way to proceed is to estimate the series under the alternative hypothesis; hence, we estimate a regression equation of the form:

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

If it is possible to reject the null hypothesis $\gamma = 0$, the process is trend stationary. The problem is to determine the appropriate value for p. Toward this end, the equation was estimated for lag lengths of 1 through 4. As given in Table 4.5, the AIC selects a lag length of three and the SBC selects a lag length of one. Nevertheless, in this instance, the lag length seems not to make a difference; at the 5% significance level, the critical value for the null hypothesis $\gamma = 0$ is -3.43. As such, the lag lengths selected by the AIC and the SBC are such that the null hypothesis of a unit root is not rejected. We can conclude that the sequence is not trend stationary.

The ϕ_3 allows us to test the null hypothesis $\gamma = a_2 = 0$; at the 5% significance level, the critical value is 6.49. As such, for any lag length, we would not reject the null hypothesis and conclude that the sequence has a stochastic trend. However, at the 5% significance level, the critical value for the null hypothesis $a_0 = \gamma = a_2 = 0$ (i.e., the critical value of the ϕ_2 statistic) is 4.88. For the lag lengths selected by the AIC and the SBC, this null hypothesis is clearly rejected. The test statistics reflect the fact that the data-generating process does contain the drift term a_0.

Table 4.5 Dickey–Fuller Tests and Lag Length

p	AIC	SBC	γ	t-Statistic	ϕ_2	ϕ_3
1	1076.211	1089.303	−0.017	−1.776	17.390	1.579
2	1073.076	1089.441	−0.020	−2.049	11.188	2.101
3	1071.817	1091.455	−0.022	−2.285	8.622	2.616
4	1073.799	1096.710	−0.022	−2.276	8.026	2.595

It is also possible to use t-tests and F-tests to determine the lag length. Estimating the equation using the lag length $p = 4$ yields

$$\begin{matrix}\Delta y_t = 1.24 + 0.042t - 0.022y_{t-1} + 0.397\Delta y_{t-1} + 0.108\Delta y_{t-2} + 0.125\Delta y_{t-3} + 0.009\Delta y_{t-4} + \varepsilon_t \\ (4.05)\ (2.28)\quad (-2.28)\quad (5.57)\quad (1.42)\quad (1.64)\quad (0.13)\end{matrix}$$

A t-test for the coefficient on Δy_{t-4} suggests a lag length no greater than 3. Moreover, the F-statistic for the null hypothesis $\beta_3 = \beta_4 = 0$ is 1.59 with a *prob*-value of 0.206. As such, we can eliminate lags 3 and 4. Moreover, the F-statistic for the null hypothesis $\beta_2 = \beta_3 = \beta_4 = 0$ is 2.76 with a *prob*-value of 0.043. Hence, if we use a 5% significance level, the F-tests select a model with two lags. In this instance, the results regarding the significance of γ are not very sensitive to the alternative lag lengths.

The standard practice is to perform your lag lengths tests first and then check for a unit root. After all, the appropriate lag length can be selected regardless of whether or not the series in question is stationary.

The Test with MA Components

Since an invertible MA model can be transformed into an autoregressive model, the procedure can be generalized to allow for moving average components. Let the $\{y_t\}$ sequence be generated from the mixed autoregressive/moving average process:

$$A(L)y_t = C(L)\varepsilon_t$$

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

If the roots of $C(L)$ are outside the unit circle, we can write the $\{y_t\}$ sequence as the autoregressive process:

$$A(L)y_t/C(L) = \varepsilon_t$$

or, defining $D(L) = A(L)/C(L)$, we can write the process as

$$D(L)y_t = \varepsilon_t$$

Even though $D(L)$ will generally be an infinite-order polynomial, in principle, we can use the same technique as used to obtain (4.30) to form the infinite-order autoregressive model:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^{\infty} \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

As it stands, this is an infinite-order autoregression that cannot be estimated using a finite data set. Fortunately, Said and Dickey (1984) have shown that an unknown ARIMA$(p, 1, q)$ process can often be well approximated by an ARIMA$(n, 1, 0)$

autoregression of order n where $n \leq T^{1/3}$. Thus, we can usually solve the problem by using a finite-order approximation of the infinite-order autoregression. The test for $\gamma = 0$ can be conducted using the aforementioned Dickey–Fuller τ, τ_μ, or τ_τ test statistics.

LAG LENGTHS AND NEGATIVE MA TERMS Unit root tests generally work poorly if the error process has a strongly negative MA component. While the result of Said and Dickey (1984) that an ARIMA$(p, 1, q)$ process can be well approximated by an ARIMA$(n, 1, 0)$ process ($n \leq T^{1/3}$), the interaction between the unit root and the negative MA component can lead to over-rejections of a unit root. To explain the nature of the problem, consider the ARIMA(0,1,1) process:

$$y_t = y_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1} \qquad 0 < \beta_1 < 1.$$

If we have the initial condition y_0, we can write the general solution for y_t as

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

Clearly, the $\{y_t\}$ sequence is not stationary since the effects of an ε_t shock never decay to zero. However, unlike a random walk process for which $\beta_1 = 0$, the presence of the negative MA term means that ε_t has a one-unit effect on y_t in period t only. Since for all subsequent periods $\partial y_{t+i}/\partial \varepsilon_t = (1 - \beta_1) < 1$, the magnitude of the effect is diminished when compared to that of a pure random walk. For a finite sample with t observations, we can construct the autocovariances as

$$\begin{aligned}
\gamma_0 &= E[(y_t - y_0)^2] = \sigma^2 + (1 - \beta_1)^2 E[(\varepsilon_{t-1})^2 + (\varepsilon_{t-2})^2 + \cdots + (\varepsilon_1)^2] \\
&= [1 + (1 - \beta_1)^2 (t - 1)]\sigma^2 \\
\gamma_s &= E[(y_t - y_0)(y_{t-s} - y_0)] \\
&= E[(\varepsilon_t + (1 - \beta_1)\varepsilon_{t-1} + \cdots + (1 - \beta_1)\varepsilon_1)(\varepsilon_{t-s} + (1 - \beta_1)\varepsilon_{t-s-1} + \cdots + (1 - \beta_1)\varepsilon_1) \\
&= (1 - \beta_1)[1 + (1 - \beta_1)(t - s - 1)]\sigma^2
\end{aligned}$$

The autocorrelations are formed from $\rho_s = \gamma_s/(\gamma_s \gamma_0)^{0.5}$. It is easy to verify that all of the autocorrelations ρ_i approach unity as the sample size t becomes infinitely large. However, for the sample sizes usually found in applied work, the autocorrelations can be small. To see the point, let β_1 be close to unity so that terms containing $(1 - \beta_1)^2$ can be safely ignored. In such circumstances, the ACF can be approximated by $\rho_1 = \rho_2 \cdots = (1 - \beta_1)^{0.5}$. For example, if $\beta_1 = 0.95$, all of the autocorrelations should be close to 0.22. As such, the autocorrelations will be small, appear to be marginally significant, and show little tendency to decay.

From the example, it should not be surprising than that unit root tests do not work well in the presence of a strongly negative MA component. Since many of the autocorrelations are small, the ACF will resemble that of a truly stationary process. In fact, if β_1 is very close to unity, there is a common factor such that $y_t = y_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1}$ approximates the white-noise process $y_t = \varepsilon_t$. Any test will have a difficult time distinguishing between the two types of processes and will over-reject the null hypothesis of a unit root. Moreover, in conducting the test, it is necessary to use a large number of lags. We can use lag operators to write $\Delta y_t = (1 - \beta_1 L)\varepsilon_t$, so that

$\Delta y_t = \beta_1 \Delta y_{t-1} + (\beta_1)^2 \Delta y_{t-2} + (\beta_1)^3 \Delta y_{t-3} + \cdots + \varepsilon_t$. When β_1 is large, many autoregressive lags are needed to properly capture the dynamics of the process. The need to estimate a large number of coefficients can diminish the power of the test.

Nevertheless, there are some precautions to take when testing for a unit root in the presence of a negative MA component. Clearly, you want to use a methodology that properly captures the need to use a large number of lags. Ng and Perron (2001) show that a modified version of the AIC (MAIC) yields a better estimate of the lag length than either the AIC or the BIC. Consider

$$\text{MAIC} = T \ln(\text{sum of squared residuals}) + 2n + 2\tau(n)$$

where $\tau(n) = \hat{\gamma}^2 \Sigma_t y_{t-1}^2 / \hat{\sigma}$, $\hat{\gamma}$ is the estimated value of γ and $\hat{\sigma}^2$ is the estimated variance.

Notice that the MAIC is equal to the usual expression for the AIC plus an additional penalty term $2\tau(n)$. Given that all models are estimated over the same sample period, the value of Σy_{t-1}^2 is the same for all models. As such, $\tau(n)$ will generally be small for models with a small value of γ^2 relative to the variance σ^2. Hence, the MAIC will tend to select the lag length resulting in a value of γ closest to that of a unit root.

At one time, it was popular to use the Phillips–Perron (1988) test if a large negative moving average component is suspected. However, the test does not generally perform as well as the Dickey–Fuller test when using the MAIC. The Phillips–Perron (1988) test is discussed in Section 4.6 of the *Supplementary Manual*.

Multiple Roots

Dickey and Pantula (1987) suggest a simple extension of the basic procedure if more than one unit root is suspected. In essence, the methodology entails nothing more than performing Dickey–Fuller tests on successive differences of $\{y_t\}$. When exactly one root is suspected, the Dickey–Fuller procedure is to estimate an equation such as $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$. In contrast, if two roots are suspected, estimate the equation:

$$\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \varepsilon_t \tag{4.32}$$

Use the appropriate statistic (i.e., τ, τ_μ, or τ_τ, depending on the deterministic elements actually included in the regression) to determine whether β_1 is significantly different from zero. If you cannot reject the null hypothesis that $\beta_1 = 0$, conclude that the $\{y_t\}$ sequence is $I(2)$. If β_1 does differ from zero, go on to determine whether there is a single unit root by estimating

$$\Delta^2 y_t = a_0 + \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \varepsilon_t \tag{4.33}$$

Since there are not two unit roots, you should find that β_1 and/or β_2 differ from zero. Under the null hypothesis of a single unit root, $\beta_1 < 0$ and $\beta_2 = 0$; under the alternative hypothesis, $\{y_t\}$ is stationary so that β_1 and β_2 are both negative. Thus, estimate (4.33) and use the Dickey–Fuller critical values to test the null hypothesis $\beta_2 = 0$. If you reject this null hypothesis, conclude that $\{y_t\}$ is stationary.

As a rule of thumb, economic series do not need to be differenced more than two times. However, in the odd case in which *at most r* unit roots are suspected, the procedure is to first estimate

$$\Delta^r y_t = a_0 + \beta_1 \Delta^{r-1} y_{t-1} + \varepsilon_t$$

If $\Delta^r y_t$ is stationary, you should find that $-2 < \beta_1 < 0$. If the Dickey–Fuller critical values for β_1 are such that it is not possible to reject the null of a unit root, you accept the hypothesis that $\{y_t\}$ contains r unit roots. If we reject this null of exactly r unit roots, the next step is to test for $r - 1$ roots by estimating

$$\Delta^r y_t = a_0 + \beta_1 \Delta^{r-1} y_{t-1} + \beta_2 \Delta^{r-2} y_{t-1} + \varepsilon_t$$

If both β_1 and β_2 differ from zero, reject the null hypothesis of $r - 1$ roots. You can use the Dickey–Fuller statistics to test the null of exactly $r - 1$ unit roots if the t-statistics for β_1 and β_2 are both statistically different from zero. If you can reject this null, the next step is to form

$$\Delta^r y_t = a_0 + \beta_1 \Delta^{r-1} y_{t-1} + \beta_2 \Delta^{r-2} y_{t-1} + \beta_3 \Delta^{r-3} y_{t-1} + \varepsilon_t$$

As long as it is possible to reject the null hypothesis that the various values of the β_i are nonzero, continue toward the equation

$$\Delta^r y_t = a_0 + \beta_1 \Delta^{r-1} y_{t-1} + \beta_2 \Delta^{r-2} y_{t-1} + \beta_3 \Delta^{r-3} y_{t-1} + \cdots + \beta_r y_{t-1} + \varepsilon_t$$

Continue in this fashion until it is not possible to reject the null of a unit root or until the $\{y_t\}$ series is shown to be stationary. Notice that this procedure is quite different from the sequential testing for successively greater numbers of unit roots. It might seem tempting to test for a single unit root, and if the null cannot be rejected, go on to test for the presence of a second root. In repeated samples, this method tends to select too few roots.

Seasonal Unit Roots

You will recall that the best-fitting model for U.S. money supply data used in Chapter 2 had the form:

$$(1 - L^4)(1 - L)(1 - a_1 L) y_t = (1 + \beta_4 L^4)\varepsilon_t$$

The specification implies that the money supply has a unit root and a seasonal unit root. Since seasonality is a key feature of many economic series, a sizable literature has been developed to test for seasonal unit roots. Before proceeding, note that the first difference of a seasonal unit root process will not be stationary. To keep matters simple, suppose that the quarterly observations of $\{y_t\}$ are generated by

$$y_t = y_{t-4} + \varepsilon_t$$

Here, the seasonal difference of $\{y_t\}$ is stationary; using the notation of Chapter 2, we can write $\Delta_4 y_t = \varepsilon_t$. Given the initial condition $y_0 = y_{-1} = \cdots = 0$, the solution for y_t is

$$y_t = \varepsilon_t + \varepsilon_{t-4} + \varepsilon_{t-8} + \cdots$$

so that

$$y_t - y_{t-1} = \sum_{i=0}^{t/4} \varepsilon_{4i} - \sum_{i=0}^{t/4} \varepsilon_{4i-1}$$

Hence, Δy_t equals the difference between two stochastic trends. Since each shock has a permanent effect on the level of Δy_t, the sequence is not mean reverting. However, the seasonal difference of a unit root process may be stationary. For example, if $\{y_t\}$ is generated by $y_t = y_{t-1} + \varepsilon_t$, the fourth difference (i.e., $\Delta_4 y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}$) is stationary. The point is that the Dickey–Fuller procedure must be modified in order to test for seasonal unit roots and distinguish between seasonal versus nonseasonal roots.

There are several alternative ways to treat seasonality in a nonstationary sequence. The most direct method occurs when the seasonal pattern is purely deterministic. For example, let D_1, D_2, and D_3 represent quarterly seasonal dummy variables such that the value of D_i is unity in season i and zero otherwise. Estimate the regression equation:

$$\Delta y_t = a_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \gamma y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \qquad (4.34)$$

The null hypothesis of a unit root (i.e., $\gamma = 0$) can be tested using the Dickey–Fuller τ_μ statistic. (Note that you use the τ_μ statistic since the original data contain an intercept). Rejecting the null hypothesis is equivalent to accepting the alternative that the $\{y_t\}$ sequence is stationary. The test is possible as Dickey, Bell, and Miller (1986) show that the limiting distribution for γ is not affected by the removal of the deterministic seasonal components. If you want to include a time trend in (4.34), use the τ_τ statistic.

Notice that the specification in (4.34) makes it difficult to test hypothesis concerning a_0. Since the mean of each D_i series is 1/4, the presence of the seasonal dummies affects the magnitude of the drift term a_0. To correct for this, it is common to use **centered** seasonal dummy variables. Simply let $D_i = 0.75$ in season i and -0.25 in each of the other three quarters of the year. Hence, the mean of $D_i = 0$ so that the magnitude of a_0 is unchanged.

If you suspect a seasonal unit root, it is necessary to use an alternative procedure. To keep the notation simple, suppose you have quarterly observations on the $\{y_t\}$ sequence and want to test for the presence of a seasonal unit root. To explain the methodology, note that the polynomial $(1 - \gamma L^4)$ can be factored such that there are four distinct characteristic roots:

$$(1 - \gamma L^4) = (1 - \gamma^{1/4} L)(1 + \gamma^{1/4} L)(1 - i\gamma^{1/4} L)(1 + i\gamma^{1/4} L) \qquad (4.35)$$

If y_t has a seasonal unit root, $\gamma = 1$. Equation (4.35) is a bit restrictive in that it only allows for a unit root at an annual frequency. Hylleberg et al. (1990) develop a clever technique that allows you to test for unit roots at various frequencies; you can test for a nonseasonal unit root, a unit root at a semiannual frequency, and/or a seasonal unit root. To understand the HEGY test (named after the four authors of the paper), suppose y_t is generated by

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

where $A(L)$ is a fourth-order polynomial such that

$$(1 - a_1L)(1 + a_2L)(1 - a_3iL)(1 + a_4iL)y_t = \varepsilon_t \tag{4.36}$$

Now, if $a_1 = a_2 = a_3 = a_4 = 1$, (4.36) is equivalent to setting $\gamma = 1$ in (4.35). Hence, if $a_1 = a_2 = a_3 = a_4 = 1$, there is a seasonal unit root. Consider some of the other possible cases:

CASE 1

If $a_1 = 1$, one homogeneous solution to (4.36) is $y_t = y_{t-1}$. As such, the $\{y_t\}$ sequence will act as a random walk in that it tends to repeat itself each and every period. This is the case of a nonseasonal unit root; the appropriate period of differencing is Δy_t.

CASE 2

If $a_2 = 1$, one homogeneous solution to (4.36) is $y_t + y_{t-1} = 0$. In this instance, the sequence tends to replicate itself at 6-month intervals so that there is a semi-annual unit root. For example, if $y_t = 1$, follows that $y_{t+1} = -1, y_{t+2} = +1, y_{t+3} = -1, y_{t+4} = 1$, and so forth.

CASE 3

If either a_3 or a_4 is equal to unity, the $\{y_t\}$ sequence has an annual cycle. For example, if $a_3 = 1$, a homogeneous solution to (4.36) is $y_t = iy_{t-1}$. Thus, if $y_t = 1$, $y_{t+1} = i$, $y_{t+2} = i^2 = -1$, $y_{t+3} = -i$, and $y_{t+4} = -i^2 = 1$ so that the sequence replicates itself every fourth period. The appropriate degree of differencing is $\Delta_4 y_t = (1 - L^4)y_t$.

To develop the test, view (4.36) as function of a_1, a_2, a_3, and a_4 and take a Taylor series approximation of $A(L)$ around the point $a_1 = a_2 = a_3 = a_4 = 1$. Although the details of the expansion are messy, first, take the partial derivative:

$$\begin{aligned}\partial A(L)/\partial a_1 &= \partial(1 - a_1L)(1 + a_2L)(1 - a_3iL)(1 + a_4iL)/\partial a_1 \\ &= -(1 + a_2L)(1 - a_3iL)(1 + a_4iL)L\end{aligned}$$

Evaluating this derivative at the point $a_1 = a_2 = a_3 = a_4 = 1$ yields

$$-L(1 + L)(1 - iL)(1 + iL) = -L(1 + L)(1 + L^2) = -L(1 + L + L^2 + L^3)$$

Next, form

$$\begin{aligned}\partial A(L)/\partial a_2 &= \partial(1 - a_1L)(1 + a_2L)(1 - a_3iL)(1 + a_4iL)/\partial a_2 \\ &= (1 - a_1L)(1 - a_3iL)(1 + a_4iL)L\end{aligned}$$

Evaluating at the point $a_1 = a_2 = a_3 = a_4 = 1$, yields $(1 - L + L^2 - L^3)L$. It should not take too long to convince yourself that evaluating $\partial A(L)/\partial a_3$ and

$\partial A(L)/\partial a_4$ at the point $a_1 = a_2 = a_3 = a_4 = 1$ yields

$$\partial A(L)/\partial a_3 = -(1 - L^2)(1 + iL)iL$$

and

$$\partial A(L)/\partial a_4 = (1 - L^2)(1 - iL)iL$$

Since $A(L)$ evaluated at $a_1 = a_2 = a_3 = a_4 = 1$ is $(1 - L^4)$, it is possible to approximate (4.36) by

$$[(1 - L^4) - L(1 + L + L^2 + L^3)(a_1 - 1) + (1 - L + L^2 - L^3)L(a_2 - 1)$$
$$- (1 - L^2)(1 + iL)iL(a_3 - 1) + (1 - L^2)(1 - iL)iL(a_4 - 1)]y_t = \varepsilon_t$$

Define γ_i such that $\gamma_i = (a_i - 1)$ and note that $(1 + iL)i = i - L$ and $(1 - iL)i = i + L$; hence,

$$(1 - L^4)y_t = \gamma_1(1 + L + L^2 + L^3)y_{t-1} - \gamma_2(1 - L + L^2 - L^3)y_{t-1}$$
$$+ (1 - L^2)[\gamma_3(i - L) - \gamma_4(i + L)]y_{t-1} + \varepsilon_t$$

so that

$$(1 - L^4)y_t = \gamma_1(1 + L + L^2 + L^3)y_{t-1} - \gamma_2(1 - L + L^2 - L^3)y_{t-1}$$
$$+ (1 - L^2)[(\gamma_3 - \gamma_4)i - (\gamma_3 + \gamma_4)L]y_{t-1} + \varepsilon_t \qquad (4.37)$$

To purge the imaginary numbers from this expression, define γ_5 and γ_6 such that $2\gamma_3 = -\gamma_6 - i\gamma_5$ and $2\gamma_4 = -\gamma_6 + i\gamma_5$. Hence, $(\gamma_3 - \gamma_4)i = \gamma_5$ and $\gamma_3 + \gamma_4 = \gamma_6$. Substituting into (4.37) yields

$$(1 - L^4)y_t = \gamma_1(1 + L + L^2 + L^3)y_{t-1} - \gamma_2(1 - L + L^2 - L^3)y_{t-1}$$
$$+(1 - L^2)(\gamma_5 - \gamma_6 L)y_{t-1} + \varepsilon_t$$

Fortunately, many software packages can perform the test directly on quarterly and monthly data. However, to understand the mechanics necessary to implement the procedure, use the following steps:

STEP 1: For quarterly data, form the following variables:

$$y_{1t-1} = (1 + L + L^2 + L^3)y_{t-1} = y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}$$
$$y_{2t-1} = (1 - L + L^2 - L^3)y_{t-1} = y_{t-1} - y_{t-2} - y_{t-3} - y_{t-4}$$
$$y_{3t-1} = (1 - L^2)y_{t-1} = y_{t-1} - y_{t-3} \qquad \text{so that } y_{3t-2} = y_{t-2} - y_{t-4}$$

STEP 2: Estimate the regression:

$$(1 - L^4)y_t = \gamma_1 y_{1t-1} - \gamma_2 y_{2t-1} + \gamma_5 y_{3t-1} - \gamma_6 y_{3t-2} + \varepsilon_t$$

You might want to modify the form of the equation by including an intercept, deterministic seasonal dummies, and a linear time trend. As in the augmented form of the Dickey–Fuller test, lagged values of $(1 - L^4)y_{t-i}$ may also be included. Perform the appropriate diagnostic checks to ensure that the residuals from the regression equation approximate a white-noise process.

STEP 3: Form the t-statistic for the null hypothesis $\gamma_1 = 0$; a selection of the appropriate critical values are reported below. If you do not reject the hypothesis $\gamma_1 = 0$, conclude that $a_1 = 1$ so that there is a nonseasonal unit root. Next, form the t-test for the hypothesis $\gamma_2 = 0$. If you do not reject the null hypothesis, conclude that $a_2 = 1$ and that there is a unit root with a semiannual frequency. Finally, perform the F-test for the hypothesis $\gamma_5 = \gamma_6 = 0$. If the calculated value is less than the critical value reported in Hylleberg et al. (1990), conclude that γ_5 and/or γ_6 is zero so that there is a seasonal unit root. Be aware that the three null hypotheses are not alternatives; a series may have nonseasonal, semiannual, and a seasonal unit roots.

At the 5% significance level, Hylleberg et al. (1990) report that the critical values are for 100 and 200 observations are

	$T = 100$			$T = 200$		
	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_5 = \gamma_6 = 0$	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_5 = \gamma_6 = 0$
Intercept	−2.88	−1.95	3.08	−2.87	−1.92	3.12
Intercept + trend	−3.47	−1.95	2.96	−3.44	−1.95	3.07
Intercept + seasonal dummies	−2.95	−2.94	6.57	−2.91	−2.89	6.62
Intercept + seasonal dummies + trend	−3.53	−2.94	6.60	−3.49	−2.91	6.57

An Example: In Chapter 2, we took the nonseasonal and the seasonal differences of the U.S. money supply and estimated a model of the form:

$$m_t = a_0 + a_1 m_{t-1} + \varepsilon_t + \beta_4 \varepsilon_{t-4}$$

where

$$m_t = (1 - L)(1 - L^4) y_t$$

and y_t is the logarithm of U.S. money supply as measured by M1.

We can use the HEGY test to determine if it is appropriate to use the seasonal and nonseasonal differences. Since it is clear that the money supply series has a sustained upward movement (see Section 11 in Chapter 2), we want to allow for the possibility that the series is TS. Hence, we include a deterministic trend and an intercept in the regression. You can open the file QUARTERLY.XLS, form y_t as above, and estimate the following regression

$$\begin{aligned}(1 - L^4) y_t = {} & \underset{(2.05)}{0.062} + \underset{(2.17)}{1.88^{*}10^{-4}} t - \underset{(-2.17)}{0.003^{*}10^{-4}} y_{1t-1} - \underset{(-4.17)}{0.668} y_{2t-1} - \underset{(-2.88)}{0.280} y_{3t-1} - \underset{(-2.24)}{0.217} y_{3t-2} \\ & + \sum_{i=1}^{3} \alpha_i D_i + \sum_{i=1}^{8} \beta_i (1 - L^4) y_{t-i}\end{aligned}$$

where the lag length of seven was chosen by the general-to-specific method beginning with a lag length of 12, the D_i are seasonal dummies, and y_{1t-1}, y_{2t-1}, y_{3t-1}, and y_{3t-2} are defined above.

The coefficient on y_{1t-1} has a t-statistic of -2.17. Given the 5% critical value, we cannot reject the null hypothesis of a nonseasonal unit root. The t-statistic for the coefficient on y_{2t-1} is -4.17; so, it is unlikely that there is a seasonal unit root at a semi annual frequency. The sample F-statistic for the null hypothesis that the coefficient on y_{3t-1} and y_{3t-2} jointly equal zero is 6.81. Hence, at the 5% significance level, there is not a seasonal unit root at the annual frequency ($6.81 < 6.57$). Thus, as in Chapter 2, it might not have been correct to difference and seasonally difference in the presence of deterministic seasonal dummy variables. As group, the seasonal dummies are highly significant; the sample F-statistic for the presence of the seasonal dummies is 7.49. Nevertheless, if you experiment with the model in the form $m_t = (1 - L)(1 - L^4)y_t$ used in Chapter 2, you should find the AR(1) and MA(4) terms perform better than a model with deterministic seasonal dummy variables. Moreover, if you perform the HEGY test without seasonal dummies, you will find both seasonal and annual unit roots.

8. STRUCTURAL CHANGE

In performing unit root tests, special care must be taken if it is suspected that structural change has occurred. When there are structural breaks, the various Dickey–Fuller test statistics are biased toward the nonrejection of a unit root. To explain, consider the situation in which there is a one-time change in the mean of an otherwise stationary sequence. In the top graph of Figure 4.9, the $\{y_t\}$ sequence was constructed so as to be stationary around a mean of zero for $t = 0, \ldots, 50$ and then to fluctuate around a mean of 6 for $t = 51, \ldots, 100$. The sequence was formed by drawing 100 normally and independently distributed values for the $\{\varepsilon_t\}$ sequence. Setting $y_0 = 0$, the next 100 values in the sequence were generated using the formula:

$$y_t = 0.5y_{t-1} + \varepsilon_t + D_L \tag{4.38}$$

where D_L is a dummy variable such that $D_L = 0$ for $t = 1, \ldots, 50$ and $D_L = 3$ for $t = 51, \ldots, 100$. The subscript L is designed to indicate that the *level* of the dummy changes. At times, it will be convenient to refer to the value of the dummy variable in period t as $D_L(t)$; in the example at hand, $D_L(50) = 0$ and $D_L(51) = 3$.

In practice, the structural change may not be as apparent as the break shown in the figure. However, the large simulated break is useful for illustrating the problem of using a Dickey–Fuller test in such circumstances. The straight line shown in the figure highlights the fact that the series appears to have a deterministic trend. In fact, the straight line is the best-fitting OLS equation:

$$y_t = a_0 + a_2 t + e_t$$

In the figure, you can see that the fitted value of a_0 is negative and the fitted value of a_2 is positive. The proper way to estimate (4.38) is to fit a simple AR(1) model and allow the intercept to change by including the dummy variable D_L. However, suppose that we unsuspectingly fit the regression equation:

$$y_t = a_0 + a_1 y_{t-1} + e_t \tag{4.39}$$

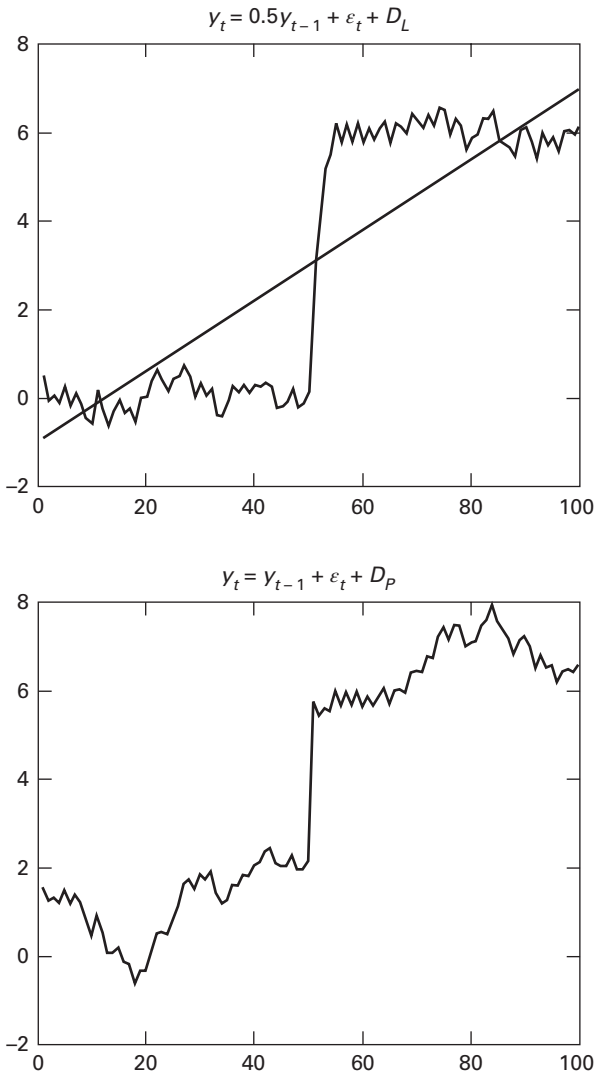


FIGURE 4.9 Two Models of Structural Change

As you can infer from Figure 4.9, the estimated value of a_1 is necessarily biased toward unity. The reason for this upward bias is that the estimated value of a_1 captures the property that "low" values of y_t (i.e., those fluctuating around zero) are followed by other "low" values, and "high" values (i.e., those fluctuating around a mean of six) are followed by other "high" values. For a formal demonstration, note that as a_1 approaches unity, (4.39) approaches a random walk plus drift. We know that the solution to the random walk plus drift model includes a deterministic trend; that is,

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

Thus, the misspecified equation (4.39) will tend to mimic the trend line shown in Figure 4.9 by biasing a_1 toward unity. This bias in a_1 means that the Dickey–Fuller test is biased toward accepting the null hypothesis of a unit root *even though the series is stationary within each of the subperiods*.

Of course, a unit root process can also exhibit a structural break. The lower portion of Figure 4.9 simulates a random walk process with a structural change occurring at $t = 51$. This second simulation used the same 100 realizations for the $\{\varepsilon_t\}$ sequence and the initial condition $y_0 = 2$. The 100 realizations of the $\{y_t\}$ sequence were constructed as

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

where $D_P(51) = 4$ and all other values of $D_P = 0$.

Here, the subscript P refers to the fact that there is a single *pulse* in the dummy variable. In a unit root process, a single pulse in the dummy will have a permanent effect on the level of the $\{y_t\}$ sequence. In $t = 51$, the pulse in the dummy is equivalent to an ε_{t+51} shock of four extra units. Hence, the *one-time* shock to $D_P(51)$ has a *permanent* effect on the mean value of the sequence for $t \geq 51$. In the figure, you can see that the level of the process takes a discrete jump in $t = 51$, never exhibiting any tendency to return to the prebreak level.

This bias in the Dickey–Fuller tests was confirmed in a Monte Carlo experiment. Perron (1989) generated 10,000 replications of a process like that of (4.38). Each replication was formed by drawing 100 normally and independently distributed values for the $\{\varepsilon_t\}$ sequence. For each of the 10,000 replicated series, he used OLS to estimate a regression in the form of (4.39). As could be anticipated from our earlier discussion, Perron found that the estimated values of a_1 were biased toward unity. Moreover, the bias became more pronounced as the magnitude of the break increased.

Perron's Test for Structural Change

Returning to the two graphs of Figure 4.9, there may be instances in which the unaided eye cannot easily detect the difference between the alternative types of sequences. One econometric procedure to test for unit roots in the presence of a structural break involves splitting the sample into two parts and using Dickey–Fuller tests on each part. The problem with this procedure is that the degrees of freedom for each of the resulting regressions are diminished. Moreover, you may not know when the breakpoint actually occurs. It is preferable to have a single test based on the full sample.

Perron (1989) goes on to develop a formal procedure to test for unit roots in the presence of a structural change at time period $t = \tau + 1$. Consider the null hypothesis of a one-time jump in the level of a unit root process against the alternative of a one-time change in the intercept of a trend stationary process. Formally, let the null and alternative hypotheses be

$$H_1 : y_t = a_0 + y_{t-1} + \mu_1 D_P + \varepsilon_t \tag{4.40}$$

$$A_1 : y_t = a_0 + a_2 t + \mu_2 D_L + \varepsilon_t \tag{4.41}$$

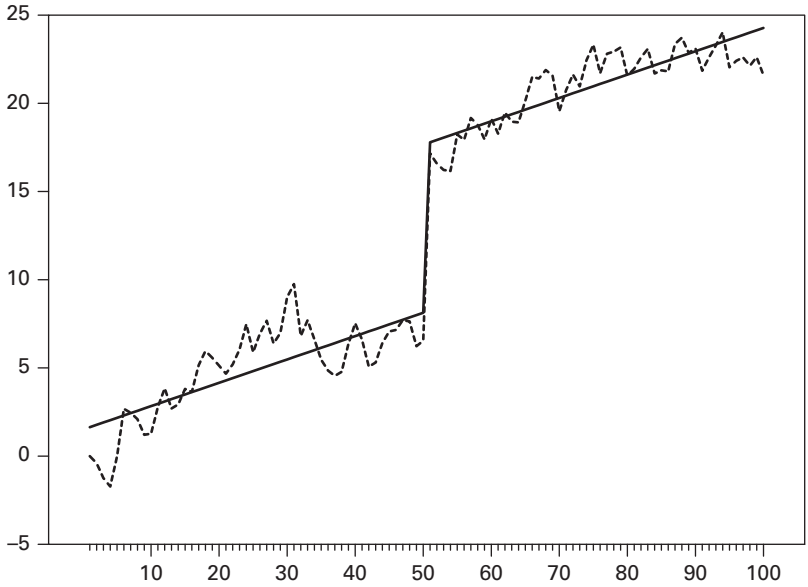


FIGURE 4.10 Alternative Representations of Structural Change

where D_P represents a *pulse* dummy variable such that $D_P = 1$ if $t = \tau + 1$ and zero otherwise and D_L represents a *level* dummy variable such that $D_L = 1$ if $t > \tau$ and zero otherwise.

Under the null hypothesis, $\{y_t\}$ is a unit root process with a one-time jump in the level of the sequence in period $t = \tau + 1$. Under the alternative hypothesis, $\{y_t\}$ is trend stationary with a one-time jump in the intercept. Figure 4.10 can help you to visualize the two hypotheses. Simulating (4.40) by setting $a_0 = 0.2$ and using 100 realizations for the $\{\varepsilon_t\}$ sequence, the erratic dashed line in the figure depicts the time path of $\{y_t\}$ under the null hypothesis. You can see the one-time jump in the level of the process occurring in period 51. Thereafter, the $\{y_t\}$ sequence continues the original random walk plus drift process. The alternative hypothesis posits that the $\{y_t\}$ sequence is stationary around the broken trend line. Up to $t = \tau$, $\{y_t\}$ is stationary around $a_0 + a_2 t$, and beginning at $\tau + 1$, y_t is stationary around $a_0 + a_2 t + \mu_2$. As illustrated by the broken line, there is a one-time increase in the intercept of the trend if $\mu_2 > 0$.

The econometric problem is to determine whether an observed series is best modeled by (4.40) or (4.41). The implementation of Perron's (1989) technique is straightforward:

STEP 1: Unlike the Dickey–Fuller test, the null hypothesis is not directly embedded in the alternative hypothesis. In other words, there is no direct way to restrict the coefficients of the alternative so as to obtain the null hypothesis. As such, we need to combine the null and alternative as follows:

$$y_t = a_0 + a_1 y_{t-1} + a_2 t + \mu_1 D_P + \mu_2 D_L + \varepsilon_t$$

STEP 2: Estimate the regression equation formed in Step 1. Under the null hypothesis of a unit root, the theoretical value of a_1 is unity. Perron (1989) shows that, when the residuals are identically and independently distributed, the distribution of a_1 depends on the proportion of observations occurring prior to the break. Denote this proportion by $\lambda = \tau/T$ where T = total number of observations.

STEP 3: Perform diagnostic checks to determine if the residuals from Step 2 are serially uncorrelated. If there is serial correlation, use the augmented form of the regression:

$$y_t = a_0 + a_1 y_{t-1} + a_2 t + \mu_1 D_P + \mu_2 D_L + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

STEP 4: Calculate the t-statistic for the null hypothesis $a_1 = 1$. This statistic can be compared to the critical values calculated by Perron. Perron generated 5000 series according to H_1 using values of λ ranging from 0 to 1 by increments of 0.1. For each value of λ, he estimated the each of the regressions and calculated the sample distribution of a_1. Naturally, the critical values are identical to the Dickey–Fuller statistics when $\lambda = 0$ and $\lambda = 1$; in effect, there is no structural change unless $0 < \lambda < 1$. The maximum difference between the two statistics occurs when $\lambda = 0.5$. For $\lambda = 0.5$, the critical value of the t-statistic at the 5% level of significance is −3.76 (which is larger in absolute than the corresponding Dickey–Fuller statistic of −3.41). If you find a t-statistic greater than the critical value calculated by Perron, it is possible to reject the null hypothesis of a unit root.

In addition, the methodology is quite general in that it can also allow for a one-time change in the drift or a one-time change in both the mean and the drift. For example, it is possible to test the null hypothesis of a permanent change in the drift term versus the alternative of a change in the slope of the trend. Here, the null hypothesis is

$$H_2\colon y_t = a_0 + y_{t-1} + \mu_2 D_L + \varepsilon_t$$

where $D_L = 1$ if $t > \tau$ and zero otherwise. With this specification, the $\{y_t\}$ sequence is generated by $\Delta y_t = a_0 + \varepsilon_t$ up to period τ and by $\Delta y_t = a_0 + \mu_2 + \varepsilon_t$ thereafter. If $\mu_2 > 0$, the magnitude of the drift increases for $t > \tau$. Similarly, a reduction in the drift occurs if $\mu_2 < 0$.

The alternative hypothesis posits a trend stationary series with a change in the slope of the trend for $t > \tau$

$$A_2\colon y_t = a_0 + a_2 t + \mu_3 D_T + \varepsilon_t$$

where $D_T = t - \tau$ for $t > \tau$ and zero otherwise.

For example, suppose that the break occurs in period 51 so that $\tau = 50$. Thus, $D_T(1)$ through $D_T(50)$ are all zero, so that, for the first 50 periods, $\{y_t\}$ evolves as $y_t = a_0 + a_2 t + \varepsilon_t$. Beginning with period 51, $D_T(51) = 1, D_T(52) = 2, \ldots$ so that, for $t > \tau$, $\{y_t\}$ evolves as $y_t = a_0 + a_2 t + \mu_3(t - 50) + \varepsilon_t = a_0 + (a_2 + \mu_3)t - 50\mu_3 + \varepsilon_t$. Hence, D_T changes the slope of the deterministic trend line. The slope of the trend is a_2 for $t \leq \tau$ and $a_2 + \mu_3$ for $t > \tau$.

To be even more general, it is possible to combine the two null hypotheses H_1 and H_2. A change in both the level and drift of a unit root process can be represented by

$$H_3: y_t = a_0 + y_{t-1} + \mu_1 D_P + \mu_2 D_L + \varepsilon_t$$

where D_P and D_L are the pulse and level dummies, respectively, defined earlier.

The appropriate alternative for this case is

$$A_3: y_t = a_0 + a_2 t + \mu_2 D_L + \mu_3 D_T + \varepsilon_t$$

Again, the procedure entails combining the null and alternative hypotheses into a single equation. Consider

$$y_t = a_0 + a_1 y_{t-1} + a_2 t + \mu_1 D_P + \mu_2 D_L + \mu_3 D_T + \varepsilon_t$$

Compare the t-statistic from the estimate of a_1 to the critical value calculated by Perron (1998). If the errors from this second regression equation do not appear to be white noise, estimate the equation in the form of an augmented Dickey–Fuller test. The t-statistic for the null hypothesis $a_1 = 1$ can be compared to the critical values calculated by Perron (1989). For various values of λ, Perron reports the following critical values of the t-statistic at the 5% significance level:

λ	H_1	H_2	H_3
0.15–0.25	−3.77	−3.80	−3.99
0.45–0.55	−3.76	−3.96	−4.24
0.65–0.75	−3.80	−3.85	−4.18

Perron's Test and Real Output

Perron (1989) used his analysis of structural change to challenge the findings of Nelson and Plosser (1982). With the same variables, his results indicate that most macroeconomic variables are *not* characterized by unit root processes. Instead, the variables appear to be TS processes coupled with structural breaks. According to Perron (1989), the stock market crash of 1929 and the dramatic oil price increase of 1973 were exogenous shocks having permanent effects on the mean of most macroeconomic variables. The crash induced a one-time fall in the mean. Otherwise, macroeconomic variables appear to be trend stationary.

All variables in the Perron's study (except real wages, stock prices, and the stationary unemployment rate) appeared to have a trend with a constant slope and exhibited a major change in the level around 1929. In order to entertain various hypotheses concerning the effects of the stock market crash, consider the regression equation:

$$y_t = a_0 + \mu_1 D_L + \mu_2 D_p + a_2 t + a_1 y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t$$

where $D_P(1930) = 1$ and zero otherwise

$D_L = 1$ for all t beginning in 1930 and zero otherwise

Table 4.6 Retesting the Data by Nelson and Plosser for Structural Change

	T	λ	k	a_0	μ_1	μ_2	a_2	a_1
Real GNP	62	0.33	8	3.44	−0.189	−0.018	0.027	0.282
				(5.07)	(−4.28)	(−0.30)	(5.05)	(−5.03)
Nominal GNP	62	0.33	8	5.69	−3.60	0.100	0.036	0.471
				(5.44)	(−4.77)	(1.09)	(5.44)	(−5.42)
Industrial production	111	0.66	8	0.120	−0.298	−0.095	0.032	0.322
				(4.37)	(−4.56)	(−0.095)	(5.42)	(−5.47)

Notes:
[1] T = number of observations; λ = proportion of observations occurring before the structural change; k = lag length.
[2] The appropriate t-statistics are in parenthesis. For a_0, μ_1, μ_2, and a_2, the null is that the coefficient is equal to zero. For a_1, the null hypothesis is $a_1 = 1$. Note that all estimated values of a_1 are significantly different from unity at the 1% level.

Under the presumption of a one-time change in the level of a unit root process, $a_1 = 1$, $a_2 = 0$, and $\mu_2 \neq 0$. Under the alternative hypothesis of a permanent one-time break in the trend stationary model, $a_1 < 1$ and $\mu_1 \neq 0$. Perron's (1989) results using real GNP, nominal GNP, and industrial production are reported in Table 4.6. Given the length of each series, the 1929 crash means that λ is 1/3 for both real and nominal GNP and equal to 2/3 for industrial production. Lag lengths (i.e., the values of k) were determined using t-tests on the coefficients β_i. The value k was selected if the t-statistic on β_k was greater than 1.60 in absolute value and the t-statistic on β_i for $i > k$ was less than 1.60.

First consider the results for real GNP. When you examine the last column of the table, it is clear that there is little support for the unit root hypothesis; the estimated value of $a_1 = 0.282$ is significantly different from unity at the 1% level. Instead, real GNP appears to have a deterministic trend (a_2 is estimated to be over five SD from zero). Also note that the point estimate $\mu_1 = -0.189$ is significantly different from zero at conventional levels. Thus, the stock market crash is estimated to have induced a permanent one-time decline in the intercept of real GNP.

These findings receive additional support since the estimated coefficients and their t-statistics are quite similar across the three equations. All values of a_1 are about five SD from unity, and the coefficients of the deterministic trends (a_2) are all over five SD from zero. Since all the estimated values of μ_1 are significant at the 1% level and negative, the data seem to support the contention that real macroeconomic variables are TS, except for a structural break resulting from the stock market crash.

Tests with Simulated Data

To further illustrate the procedure, 100 random numbers were drawn to represent the $\{\varepsilon_t\}$ sequence. By setting $y_0 = 0$, the next 100 values in the $\{y_t\}$ sequence were drawn as

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

where $D_L = 0$ for $t = 1, \ldots, 50$ and $D_L = 1$ for $t = 51, \ldots, 100$

Thus, the simulation is identical to (4.38) except that the magnitude of the structural break is diminished. This simulated series is in the data file labeled BREAK.XLS; you should try to reproduce the following results. If you were to plot the data, you would see the same pattern as in Figure 4.10. However, if you did not plot the data or were otherwise unaware of the break, you might easily conclude that the $\{y_t\}$ sequence had a unit root. The ACF of the $\{y_t\}$ sequence suggests a unit root process; for example, the first six autocorrelations are

	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
Levels	0.95	0.89	0.86	0.84	0.80	0.77
First differences	−0.002	−0.211	−0.112	0.083	−0.007	−0.025

Dickey–Fuller tests yield

$$\Delta y_t = -0.0233 y_{t-1} + \varepsilon_t \qquad t\text{-statistic for } \gamma = 0\text{: } -0.985$$
$$\Delta y_t = 0.0661 - 0.0566 y_{t-1} + \varepsilon_t \qquad t\text{-statistic for } \gamma = 0\text{: } -1.706$$
$$\Delta y_t = -0.0488 - 0.1522 y_{t-1} + 0.004t + \varepsilon_t \qquad t\text{-statistic for } \gamma = 0\text{: } -2.734$$

Diagnostic tests indicate that longer lags are not needed. Regardless of the presence of the constant or the trend, the $\{y_t\}$ sequence appears to be DS. Of course, the problem is that the structural break biases the data toward suggesting a unit root.

Now, using the Perron procedure, the first step is to estimate the model

$$\underset{}{y_t} = \underset{(1.30)}{0.083} + \underset{(5.52)}{0.479 y_{t-1}} - \underset{(-1.25)}{0.002t} + \underset{(0.076)}{0.025 D_P} + \underset{(5.52)}{0.479 D_L} + \varepsilon_t$$

In the next step, all of the diagnostic statistics indicate that $\{\varepsilon_t\}$ approximates a white-noise process. Finally, since the standard error of a_1 is 0.0897, the t-statistic for $a_1 = 1$ is −6.01 (i.e., a_1 is about six SD from unity). Since the 5% critical value is −3.76, we can reject the null of a unit root and conclude that the simulated data are stationary around a break point at $t = 51$.

Some care must be exercised in using Perron's procedure since it assumes that the date of the structural break is known. In your own work, if the date of the break is uncertain, you should consult Amsler and Lee (1995), Perron (1997), Vogelsang and Perron (1998), Zivot and Andrews (1992), Enders and Lee (2012), or Lee and Strazicich (2003). The entire issue of the July 1992 *Journal of Business and Economic Statistics* is devoted to breakpoints and unit roots. An interesting application is found in Ben-David and Papell (1995). They consider a long span (of up to 130 years) of GDP data for 16 countries. Allowing for breaks, they reject the null of a unit root in approximately half of the cases. The appropriate use of the tests of Perron (1989), Zivot and Andrews (1992), and Lee and Strazicich (2003) are shown in Chapter 6 of the *Programming Manual*.

9. POWER AND THE DETERMINISTIC REGRESSORS

Tests for unit roots are not especially good at distinguishing between a series with a characteristic root that is close to unity and a true unit root process. Part of the problem concerns the power of the test and the presence of the deterministic regressors in the estimating equations.

Power

Formally, the **power** of a test is equal to the probability of rejecting a false null hypothesis (i.e., one minus the probability of a type II error). A test with good power would correctly reject the null hypothesis of a unit root when the series in question is actually stationary. Monte Carlo simulations have shown that the power of the various Dickey–Fuller tests can be very low. As such, these tests will too often indicate that a series contains a unit root. The problem is that, in finite samples, any trend stationary process can be arbitrarily well approximated by a unit root process, and a unit root process can be arbitrarily well approximated by a trend stationary process. To explain, examine the interest rate series and exchange rate series shown in the beginning of Chapter 3. If you did not know the actual data-generating processes, it would be difficult to tell which, if any, are stationary. Similarly, it is difficult for any statistical procedure to distinguish between unit root processes and series that are highly persistent.

It is simple to conduct a Monte Carlo experiment that determines the power of the Dickey–Fuller test. To be more specific, suppose that the true data-generating process for a series is $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$ where $|a_1| < 1$. If you did not know the actual data-generating process, you might test the series for a unit root using a Dickey–Fuller test. The question at hand is How often will the Dickey–Fuller test fail to detect that the series is actually stationary? Since the confidence intervals for the t-statistics of the Dickey–Fuller exceed those for the usual t-test, it is to be expected that the power of the Dickey–Fuller test is low. To find out the exact answer to the question, we can generate 10,000 stationary series and apply a Dickey–Fuller test to each. We can then calculate the percentage of the times that the test correctly identifies a truly stationary process.

The ability of the test to properly detect that the series is stationary will depend on the value of a_1. We would expect the test to have the least power when $|a_1|$ is close to unity. Thus, it makes sense to examine how the magnitude of a_1 affects the power of the test. We first construct 100 observations of the series $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$ using a value of $a_1 = 0.8$ and an $\{\varepsilon_t\}$ sequence drawn from a standardized normal distribution. The magnitude of a_0 is unimportant and, so, is set equal to zero. The initial value of y_0 is set equal to the unconditional mean of zero. Next, the simulated series is estimated in the form $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$. The Dickey–Fuller τ_μ statistics are used to determine whether the null hypothesis that $\gamma = 0$ can be rejected at the 10%, 5%, and 1% significance levels. The experiment is repeated 10,000 times, and the proportion of the instances in which the null hypothesis is correctly rejected is recorded. Finally,

the entire experiment is repeated for other values of a_1. Consider the following table of proportions:

a_1	10%	5%	1%
0.80	95.9	87.4	51.4
0.90	52.1	33.1	9.0
0.95	23.4	12.7	2.6
0.99	10.5	5.8	1.3

When the true value of $a_1 = 0.8$, the test does reasonably well. For example, at the 5% significance level, the false null hypothesis of a unit root is rejected in 87.4% of the Monte Carlo replications. However, when $a_1 = 0.95$, the probability of correctly rejecting the null hypothesis of a unit root is estimated to be only 12.7% at the 5% significance level and 2.6% at the 1% level. Thus, the test has very low power to detect near unit root series.

Does it matter that it is often impossible to distinguish between borderline stationary, trend stationary, and unit root processes? The realistic answer is that it depends on the question at hand. In borderline cases, the short-run forecasts from the alternative models may have nearly identical forecasting performance. In fact, Monte Carlo studies indicate that when the true data-generating process is stationary but has a root close to unity, the one-step-ahead forecasts from a differenced model are usually superior to the forecasts from a stationary model. However, the long-run forecasts of a model with a deterministic trend will be quite different from those of other models.

Determination of the Deterministic Regressors

Unless the researcher knows the actual data-generating process, there is a question concerning whether it is most appropriate to estimate (4.20), (4.21), or (4.22). It might seem reasonable to test the hypothesis $\gamma = 0$ using the most general of the models:

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t \tag{4.44}$$

After all, if the true process is a random walk process, this regression should find that $a_0 = \gamma = a_2 = 0$. One problem with this line of reasoning is that the presence of the additional estimated parameters reduces degrees of freedom and the power of the test. Reduced power means that the researcher will not be able to reject the null of a unit root when, in fact, no unit root is present. The second problem is that the appropriate statistic (i.e., τ, τ_μ, and τ_τ) for testing $\gamma = 0$ depends on which regressors are included in the model. As you can see by examining the three Dickey–Fuller tables, for a given significance level, the confidence intervals around $\gamma = 0$ dramatically expand if a drift and a time trend are included in the model. This is quite different from the case in which $\{y_t\}$ is stationary. When $\{y_t\}$ is stationary, the distribution of the t-statistic does not depend on the presence of other regressors.

The point is that it is important to use a regression equation that mimics the actual data-generating process. Inappropriately omitting the intercept or time trend can cause the power of the test to go to zero. For example, if as, in (4.44), the data-generating process includes a trend, omitting the term $a_2 t$ imparts an upward bias in the estimated value of γ. On the other hand, extra regressors increase the critical values so that you may fail to reject the null of a unit root.

Campbell and Perron (1991) report the following results concerning unit root tests:

1. If the estimated regression includes deterministic regressors that are not in the actual data-generating process, the power of the unit root test against a stationary alternative decreases as additional deterministic regressors are added. Hence, you do not want to include regressors that are not in the data-generating process.
2. If the estimated regression omits an important deterministic trending variable present in the true data-generating process—such as the expression $a_2 t$ in (4.44)—the power of the t-statistic test goes to zero as sample size increases. If the estimated regression omits a nontrending variable (such as an intercept), the t-statistic is consistent, but the finite sample power is adversely affected and decreases as the magnitude of the coefficient on the omitted component increases. Hence, you do not want to omit regressors that are in the data-generating process.

The direct implication of these findings is that the researcher may fail to reject the null hypothesis of a unit root because of a misspecification concerning the deterministic part of the regression. Too few or too many regressors may cause a failure of the test to reject the null of a unit root. How do you know whether to include a drift or time trend in performing the tests? The key problem is that *the tests for unit roots are conditional on the presence of the deterministic regressors and tests for the presence of the deterministic regressors are conditional on the presence of a unit root*. Although we can never be sure that we are including the appropriate deterministic regressors in our econometric model, there are some useful guidelines.

1. *Always plot your data.* Visual inspection can help you determine whether there is a clear trend in the data.
2. *Be clear about the appropriate null hypothesis and the alternative hypothesis*. When you perform a Dickey–Fuller test, always estimate the model under the alternative hypothesis and impose the restriction implied by the null hypothesis. Since the null hypothesis is that the series has a unit root, always estimate the series as if it were stationary or TS. For example, the real GDP series shown in Figure 4.1 moves decidedly upward over time. The issue is whether the series is trend stationary or contains a unit root plus a drift term. As such, the appropriate model to estimate has the form $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \Sigma \beta_i \Delta y_{t-i} + \varepsilon_t$. You then test the restrictions $\gamma = 0$ and/or $\gamma = a_2 = 0$. There is no need to estimate a model without $a_2 t$ since the alternative hypothesis is not represented in such a specification.

3. *You do not want to reject the null hypothesis when the series actually has a unit root (a Type I error) or incorrectly accept the null of a unit root when the series is stationary or TS (a Type II error)*. Nevertheless, any test involves the possibility of making such errors. As such, you do not want to perform needless tests. In the example of real GDP, there is little point in testing the restriction $a_0 = \gamma = a_2 = 0$ since real GDP clearly increases over time.
4. *Testing a restriction on a model that has already been restricted creates the possibility of compounding your errors*. Suppose that a test for the presence of the time trend allows you to set $a_2 = 0$. A subsequent test of the restriction $a_0 = \gamma = 0$ in the model $\Delta y_t = a_0 + \gamma y_{t-1} + \Sigma\beta_i \Delta y_{t-i} + \varepsilon_t$ is conditional on whether the first test was correct in allowing you to exclude the deterministic trend.

At one time, researchers would apply a battery of tests on the values of a_0 and/or a_2 when the form of the deterministic regressors was completely unknown. One standard procedure is discussed in Section 4.4 of the *Supplementary Manual* and in Chapter 6 of the *Programming Manual*. Now, when power seems to be an issue, it is typical to use variants of the Dickey–Fuller test that have enhanced power.

10. TESTS WITH MORE POWER

If you examine the basic regression used in the Dickey–Fuller test, $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \varepsilon_t$, you will see that there are two different types of regressors. The intercept and the time trend are purely deterministic while y_{t-1} is a unit root process under the null hypothesis. Notice that the coefficients of the deterministic expressions, a_0 and a_2, play very different roles under the null and alternative hypotheses. If we change equation numbers and symbols to match those used in the text, Phillips and Schmidt (1992, p. 258) make the following observation about the parameters in the Dickey–Fuller regressions

> "… the parameter a_0 represents trend when $\gamma = 0$ (since the solution for y_t, then includes the deterministic trend term $a_0 t$), but it determines level when $\gamma < 0$ (since y_t is then stationary around the level $-a_0/\gamma$). Similarly, [in equation (4.44)], when $\gamma = 0$, the parameter a_0 represents trend and a_2 represents quadratic trend, while under the alterative a_0 determines level and a_2 determines trend. This confusion over the meanings of the parameters shows up in the properties of the Dickey–Fuller tests."

The essential problem is that the intercept and the slope of the trend are often poorly estimated in the presence of a unit root. In a sense, the least squares principle is unable to properly separate the movements of y_t into those induced by the deterministic trend and those induced by the stochastic trend. Even in the circumstance in which $\{y_t\}$ is stationary, the intercept and trend can be poorly estimated if the $\{y_t\}$ series is quite persistent. Of course, if the estimates of a_0 and a_2 have substantial error, the estimate of γ will have a large standard error too. You can see this effect by comparing the Dickey–Fuller critical values for τ, τ_μ, and τ_τ to those in a standard t-table. The overly

wide confidence interval for γ means that you are less likely to reject the null hypothesis of a unit root even when the true value of γ is not zero.

A number of authors have devised clever methods to improve the estimates of the intercept and trend coefficients. For example, Schmidt and Phillips (1992) proposed a two-step testing procedure that has better power than the Dickey–Fuller test. Although they call their test a **Lagrange Multiplier** (LM) test, the method is actually quite simple. Instead of the Dickey–Fuller specification, under the null hypothesis, the $\{y_t\}$ sequence is a random walk plus a drift so that:

$$y_t = a_0 + a_2 t + \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

or

$$\Delta y_t = a_2 + \varepsilon_t$$

The idea is to estimate the trend coefficient, a_2, using the regression $\Delta y_t = a_2 + \varepsilon_t$. As such, the presence of the stochastic trend $\Sigma\varepsilon_i$ does not interfere with the estimation of a_2. The resulting estimate of a_2 (called $\hat{a}_2$) is an estimate of the slope of the time trend. Use this estimate to form the detrended series as $y_t^d = y_t - (y_1 - \hat{a}_2) - \hat{a}_2 t$, where y_1 is the initial value of the $\{y_t\}$ series. Note that $(y_1 - \hat{a}_2)$ acts as the intercept of the estimated trend line and $\hat{a}_2$ acts as the slope. The use of $(y_1 - \hat{a}_2)$ in the detrending procedure ensures that the initial value of the detrended series (i.e., y_1^d) is zero. In the second step of the procedure, you estimate a variant of the Dickey–Fuller test using the detrended series in place of the level of y_{t-1}

$$\Delta y_t = a_0 + \gamma y_{t-1}^d + \varepsilon_t$$

or, if there is any serial correlation in the residuals, estimate

$$\Delta y_t = a_0 + \gamma y_{t-1}^d + \sum_{i=1}^{p} c_i \Delta y_{t-i}^d + \varepsilon_t$$

The null of a unit root can be rejected if it is found that $\gamma \neq 0$. The point is that Schmidt and Phillips (1992) show that it is preferable to estimate the parameters of the trend using a model without the persistent variable y_{t-1}. Once the trend is efficiently estimated, it is possible to detrend the data and perform the unit root test on the detrended data. Some of the critical values for the test are

Critical Values of the Schmidt–Phillips Unit Root Test

T	1%	2.5%	5%	10%
50	−3.73	−3.39	−3.11	−2.80
100	−3.63	−3.32	−3.06	−2.77
200	−3.61	−3.30	−3.04	−2.76
500	−3.59	−3.29	−3.04	−2.76

Elliott, Rothenberg, and Stock (1996) show that it is possible to further enhance the power of the test by estimating the model using something close to first differences. The idea is that, under the alternative hypothesis that the series is stationary, the Schmidt–Phillips model in first differences is misspecified. Hence, consider the TS model:

$$y_t = a_0 + a_2 t + B(L)\varepsilon_t$$

Instead of creating the first difference of y_t, Elliott, Rothenberg, and Stock (ERS) preselect a constant close to unity, say α, and subtract αy_{t-1} from y_t to obtain

$$\widetilde{y}_t = a_0 + a_2 t - \alpha a_0 - \alpha a_2 (t-1) + e_t, \qquad \text{for } t = 2, \ldots, T$$

where $\widetilde{y}_t = y_t - \alpha y_{t-1}$ and e_t is a stationary error term. For $t = 1$, such near differencing is not possible and the initial value $\widetilde{y}_1$ is set equal to y_1. For simplicity, collect terms with a_0 and a_2 to obtain

$$\widetilde{y}_t = (1-\alpha)a_0 + a_2[(1-\alpha)t + \alpha)] + e_t$$

Now, it should be clear how to obtain estimates of a_0 and a_2 using OLS. Create the variable $z1_t$ equal to the constant $(1-\alpha)$ and the variable $z2_t$ equal to the deterministic trend $\alpha + (1-\alpha)t$. To obtain the desired estimates of a_0 and a_2, simply regress $z1_t$ and $z2_t$ on $\widetilde{y}_t$. In other words, use OLS to estimate:

$$\widetilde{y}_t = a_0 z1_t + a_2 z2_t + e_t$$

Note that the test is conditional on the initial value of the $\{y_t\}$ series in that $y_1 = a_0 + a_2 + \varepsilon_1$. As such, the initial values of $z1_t$ and $z2_t$ should be set equal to unity and the initial value of $\widetilde{y}_t$ should be set equal to y_1 (i.e., set $z1_1 = 1$, $z2_1 = 1$, and $\widetilde{y}_1 = y_1$). Since the goal is to obtain the estimated values of a_0 and a_2, at this step, it is not especially important if the residual, e_t, is serially correlated. The important point is that the estimates a_0 and a_2 can be used to detrend the $\{y_t\}$ series as

$$y_t^d = y_t - \hat{a}_0 - \hat{a}_2 t$$

In the second step of the procedure, estimate the basic Dickey–Fuller regression using the detrended data. Hence, estimate the regression equation:

$$\Delta y_t^d = \gamma y_{t-1}^d + \varepsilon_t$$

If there is serial correlation in the residuals, the augmented form of the test can be estimated as

$$\Delta y_t^d = \gamma y_{t-1}^d + \sum_{i=1}^{p} c_i \Delta y_{t-i}^d + \varepsilon_t$$

Elliott, Rothenberg, and Stock (1996) recommend selecting the lag length p using the SBC. As in the Schmidt–Phillips test, the null of a unit root can be rejected if it is found the $\gamma \neq 0$. The critical values of the test depend on whether a trend is included in the test. *If there is an intercept but not a trend, the critical values are precisely those of*

the Dickey–Fuller τ test reported in the top portion of Table A. In essence, you use the Dickey–Fuller critical values as if there is no intercept in the data-generating process. If there is a trend, the critical values depend on the value of α selected to perform the "near differenced" variable $\hat{y}_t$. ERS report that the value of α that seems to provide the best overall power is $\alpha = (1 - 7/T)$ for the case of an intercept and $\alpha = (1 - 13.5/T)$ if there is an intercept and trend. The table below reports the critical values for the case of a trend and $\alpha = 1 - 13.5/T$. Notice that, as the sample size T increases, α approaches unity so that $\hat{y}_t$ is approximately equal to Δy_t. In the literature, the ERS test is often referred to as the Dickey–Fuller generalized least squares (**DF-GLS**) test.[5]

Critical Values of the ERS Test with Trend and $\alpha = 1 - 13.5 / T$

T	1%	2.5%	5%	10%
50	−3.77	−3.46	−3.19	−2.89
100	−3.58	−3.29	−3.03	−2.74
200	−3.46	−3.18	−2.93	−2.64
∞	−3.48	−3.15	−2.89	−2.57

One aspect of the ERS test that some researchers might find objectionable is the assumption that the initial value $\tilde{y}_1$ is set equal to y_1. This is equivalent to assuming that the initial value of the error term is equal to zero. An alternative assumption is that the initial value of the shock is drawn from its unconditional distribution. Note that relaxing the assumption concerning the initial condition acts to reduce the power of this version of the test. In this circumstance, the first value of $\tilde{y}_1$ is set equal to $y_1(1 - \alpha^2)^{0.5}$, $z1_1 = (1 - \alpha^2)^{0.5}$, and $z2_1 = (1 - \alpha^2)^{0.5}$.[6] Hence, instead of conditioning on the magnitude of y_1, you condition on the number of SD from zero. Note that Elliott (1999) recommends using $\alpha = (1 - 10/T)$ regardless of whether or not a trend is included in the regression. The critical values for this test are different from those reported above. The asymptotic critical values for regressions with an intercept and an intercept plus trend are as follows:

	1%	2.5%	5%	10%
Intercept	−3.28	−2.98	−2.73	−2.46
Trend	−3.71	−3.41	−3.17	−2.91

An Example

In order to illustrate the appropriate use of the procedure, the file labeled ERSTEST.XLS contains 200 observations generated from the equation: $y_t = 1 + 0.95y_{t-1} + 0.01t + \varepsilon_t$. Although the series is clearly trend stationary, the point of this exercise is to illustrate the appropriate use of the ERS test and compare the results

to those of a Dickey–Fuller test. If you examine the file, you will see that the first five rows are

t	*y*	*y_tilde*	*z1*	*z2*	*yd*
1	20.03339	20.03339	1.0000	1.0000	0.036376
2	21.85126	3.170125	0.0675	1.0675	1.692188
3	22.01347	1.637169	0.0675	1.1350	1.692338
4	22.08649	1.558934	0.0675	1.2025	1.603304
5	22.17255	1.576890	0.0675	1.2700	1.527297

The series in column 2, called y, contains the 200 realizations representing the y_t series. Since the data contain a trend, the appropriate value of α to use is $1 - 13.5/200 = 0.9325$. This value of α was used to construct the next series (*y_tilde*) as $y_t - 0.9325y_{t-1}$. For example, $\tilde{y}_1 = y_1$, $\tilde{y}_2 = y_2 - \alpha y_1 = 21.85126 - 0.9325(20.03339) = 3.170125$ and $\tilde{y}_3 = y_3 - \alpha y_2 = 1.637169$. Since $\alpha = 0.9325$, $z1_2 = z1_3 = \cdots = 1 - \alpha = 0.0675$. Similarly, $z2_t = 0.9325 + 0.0675t$ so that $z2_1 = 1.0000$, $z2_2 = 1.0675$, $z2_3 = 1.1350$, ... The regression of $\tilde{y}_t$ on $z1_t$ and $z2_t$ yields

$$\tilde{y}_t = 19.835 * z1_t + 0.162 * z2_t$$

These estimates of a_0 and a_2 are used to construct the detrended series as

$$y_t^d = y_t - 19.835 - 0.162t$$

This series is reported in the last column of ERSTEST.XLS. Before proceeding, it is interesting to consider the particular solution for the skeleton of $y_t = 1 + 0.95y_{t-1} + 0.01t + \varepsilon_t$. From your knowledge of Chapter 1 (also see question 7 of Chapter 2), you should have no trouble verifying that the desired solution is $16.2 + 0.2t$. The estimated trend equation, $19.835 + 0.162t$, is reasonably close to the particular solution.

Now that y_t has been detrended, it is straightforward to perform the unit root test. If you use the data in the spreadsheet, you should find

$$\Delta y_t^d = -0.0975y_{t-1}^d$$
$$(-3.154)$$

The 2.5% and 5% critical values for the test are −3.15 and −2.89, respectively. As such, the null hypothesis of a unit root is clearly rejected at the 5% level and just barely rejected at the 2.5% level. You will find that augmenting this regression with lagged values of Δy_{t-i}^d only acts to increase the value of the SBC. You can perform Elliott's (1999) version of the test in the same way, except that you set $\alpha = 1 - 10/200 = 0.95$, $y_1(1 - \alpha^2)^{0.5} = 6.255$, $z1_1 = (1 - \alpha^2)^{0.5} = 0.3122$, and $z2_1 = (1 - \alpha^2)^{0.5} = 0.3122$. Hence, assuming that the initial value of the series is drawn from its unconditional mean, you should obtain the t-statistic −3.147. The null hypothesis of a unit root is not rejected (although it is very close to being rejected) using the 5% critical value of −3.17.

The results of Elliott's (1999) test are very similar to the result found from the Schmidt–Phillips test. To perform the Schmidt–Phillips LM test, you should first regress Δy_t on a constant and obtain: $\Delta y_t = 0.1713$. Since $y_1 = 20.03339$, you detrend

the y_t series using $y_t^d = 20.03339 - (20.03339 - 0.1713) - 0.1713t$. Now, you should be able to reproduce the regression equation $\Delta y_t = 0.0691 - 0.0903 y_t^d$. Since the t-statistic for the coefficient on y_t^d is -3.052, the null hypothesis of a unit root is just rejected at the 5% significance level. Very different results are obtained when performing a standard Dickey–Fuller test. Consider the estimated model:

$$\Delta y_t = 2.0809 + 0.0158t - 0.0979 y_{t-1} + \varepsilon_t$$
$$(3.265) \quad (3.106) \quad (-3.124)$$

The estimated value of γ is -0.0979, and the t-statistic for the null hypothesis $\hat{\gamma} = 0$ is -3.124. From Table A, the critical values of the τ_τ statistic at the 5% and 10% significance levels are about -3.45 and -3.15, respectively. Hence, if we use the Dickey–Fuller test, the null hypothesis of a unit root cannot be rejected at conventional significance levels.

Section 9 reported the results of a simple Monte Carlo study of the power of the standard Dickey–Fuller test for the process: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$. Now, if we use the ERS test, the proportions (out of 10,000 replications) in which the null hypothesis of a unit root were correctly rejected are

a_1	10%	5%	1%
0.80	99.8	99.1	86.6
0.90	93.9	79.0	33.4
0.95	64.3	39.8	10.0
0.99	23.3	11.1	2.3

Although these results are far superior to those of the Dickey–Fuller test, the power of the test for large values of a_1 is still disappointing.

Section 6.3 of the *Programming Manual* uses real U.S. DGP to illustrate the appropriate use of the test.

11. PANEL UNIT ROOT TESTS

Section 6 presented some strong evidence that the three real exchange rate series shown in Figure 4.7 are unit root processes. Of course, it is possible that the series are mean reverting but the Dickey–Fuller tests have little power to detect the fact that the series are stationary. One way to obtain a more powerful test is to pool the estimates from a number of separate series and then test the pooled value. The theory underlying the test is very simple: if you have n independent and unbiased estimates of a parameter, the mean of the estimates is also unbiased. More importantly, so long as the estimates are independent, the central limit theory suggests that the sample mean will be normally distributed around the true mean.

Im, Pesaran, and Shin (2002) show how to use this result to construct a test for a unit root when you have a number of similar time-series variables (i.e., a panel). The only complicating factor is that the OLS estimates for γ in the Dickey–Fuller test are biased downward. Suppose you have n series each containing T observations. For each

series, perform an ADF test of the form

$$\Delta y_{it} = a_{i0} + \gamma_i y_{it-1} + a_{i2} t + \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{it-j} + \varepsilon_{it} \qquad i = 1, \ldots, n \tag{4.45}$$

Because the lag lengths can differ across equations, you should perform separate lag length tests for each equation. Moreover, you may choose to exclude the deterministic time trend. However, if the trend is included in one equation, it should be included in all. Once you have estimated the various γ_i, obtain the t-statistic for the null hypothesis $\gamma_i = 0$. In a traditional Dickey–Fuller test, each of these t-statistics—denoted by t_i—would be compared to the appropriate critical value reported in Table A. However, for the panel unit root test, form the sample mean of the t-statistics as

$$\bar{t} = (1/n) \sum_{i=1}^{n} t_i \tag{4.46}$$

It is straightforward to construct the statistic $Z_{\bar{t}}$ as

$$Z_{\bar{t}} = \frac{\sqrt{n}[\bar{t} - E(\bar{t})]}{\sqrt{\text{var}(\bar{t})}}$$

where $E\bar{t}$ and $\text{var}(\bar{t})$ denote the theoretical mean and variance of $\bar{t}$. If the OLS estimates of the individual t_i were unbiased, the value of $E\bar{t}$ would be zero. However, to correct for the bias, the values $E\bar{t}$ and $\text{var}(\bar{t})$ can be calculated by Monte Carlo simulation. Im, Pesaran, and Shin (IPS) report these values as follows:

T	6	8	10	15	20	50	100	500
$E\bar{t}$	−1.52	−1.50	−1.50	−1.51	−1.52	−1.53	−1.53	−1.53
$\text{var}(\bar{t})$	1.75	1.23	1.07	0.92	0.85	0.76	0.74	0.72

Im, Pesaran, and Shin show that $Z_{\bar{t}}$ has an asymptotic standardized normal distribution. Hence, for large T and n, you can approximate $Z_{\bar{t}}$ with a normal distribution. This fact should not be too surprising. If each of the estimated values of the various t_i are independent, the central limit theorem indicates that deviation of the sample average from the true mean will have a normal distribution. Rejecting the null hypothesis $Z_{\bar{t}} = 0$ is equivalent to accepting the alternative hypothesis that *at least* one value of the γ_i differs from zero. After all, if the sample average of the t-statistics is significantly different from zero, at least one of the values of γ_i is statistically different from zero.

The proof that $Z_{\bar{t}}$ has a normal distribution relies on very large samples. For the sample sizes typically used by applied econometricians, it is preferable to use the critical values contained in Table 4.7. Notice that the critical values depend on n, T, and whether a time trend is included in (4.45). For example, if you have seven series each containing 50 observations and you include a time trend in (4.45), the 5% critical value for $\bar{t}$ is −2.67. If you had used the Dickey–Fuller test, the corresponding critical value for each of the seven values of t_i would be −3.50 (see Table A). Note that it is necessary to have values of T and n, which are greater than four. Large values of T are standard

Table 4.7 Selected Critical Values for the IPS Panel Unit Root Test

	25			50			70		
n/T	10%	5%	1%	10%	5%	1%	10%	5%	1%
				No Time Trend					
5	−2.04	−2.18	−2.46	−2.02	−2.15	−2.42	−2.02	−2.15	−2.40
7	−1.95	−2.08	−2.32	−1.95	−2.06	−2.28	−1.95	−2.06	−2.28
10	−1.88	−1.99	−2.19	−1.88	−1.98	−2.16	−1.88	−1.98	−2.16
15	−1.82	−1.90	−2.07	−1.81	−1.89	−2.05	−1.81	−1.89	−2.04
25	−1.75	−1.82	−1.94	−1.75	−1.81	−1.93	−1.75	−1.81	−1.93
50	−1.69	−1.73	−1.82	−1.68	−1.73	−1.81	−1.64	−1.67	−1.73
				Time Trend					
5	−2.65	−2.80	−3.09	−2.62	−2.76	−3.02	−2.62	−2.75	−3.00
7	−2.58	−2.70	−2.94	−2.56	−2.67	−2.88	−2.55	−2.66	−2.67
10	−2.51	−2.62	−2.82	−2.50	−2.59	−2.77	−2.49	−2.58	−2.75
15	−2.45	−2.53	−2.69	−2.44	−2.52	−2.65	−2.44	−2.51	−2.65
25	−2.39	−2.45	−2.58	−2.38	−2.44	−2.55	−2.38	−2.44	−2.54
50	−2.33	−2.37	−2.45	−2.32	−2.36	−2.44	−2.32	−2.36	−2.44

in time-series econometrics. However, if n is too small, the calculation of $\bar{t}$ will not be meaningful.

As mentioned in Section 6, the file PANEL.XLS contains quarterly values of the real effective exchange rates (CPI based) for Australia, Canada, France, Germany, Japan, the Netherlands, the United Kingdom, and the United States over the $1980Q1-2013Q1$ period. Since PPP does not allow for a deterministic time trend, each was estimated in the form of (4.45) but without the trend. The results of the individual Dickey–Fuller tests for the logarithmic values of the real rates are shown in the first four columns of Table 4.8. For example, the Australian equation used five lags of $\{\Delta y_{it}\}$ and the estimated value of γ_i was -0.049. Notice that the eight t-statistics for the null hypothesis $\gamma_i = 0$ have an average value of -2.44. Since each series has a total of 133 observations, the critical values at the 5% and 1% significance levels are

Table 4.8 The Panel Unit Root Tests for Real Exchange Rates

	Lags	Estimated γ_i	t-Statistic	Estimated γ_i	t-Statistic
		Log of the Real Rate		**Minus the Common Time Effect**	
Australia	5	−0.049	−1.678	−0.043	−1.434
Canada	7	−0.036	−1.896	−0.035	−1.820
France	1	−0.079	−2.999	−0.102	−3.433
Germany	1	−0.068	−2.669	−0.067	−2.669
Japan	3	−0.054	−2.277	−0.048	−2.137
The Netherlands	1	−0.110	−3.473	−0.137	−3.953
The United Kingdom	1	−0.081	−2.759	−0.069	−2.504
The United States	1	−0.037	−1.764	−0.045	−2.008

about −2.06 and −2.28, respectively. Hence, it is possible to reject the null hypothesis that all values of $\gamma_i = 0$.

One problem with the results is that the residuals from the individual equation are contemporaneously correlated in that $E\varepsilon_{it}\varepsilon_{jt} \neq 0$. For example, the correlation coefficient between the residuals from the French and German equations is 0.67. The explanation is that the shocks that affect the French real rate are likely to affect the German real rate. In this circumstance, a common strategy is to subtract a common time effect from each observation. At time period t, the mean value of each series is

$$\overline{y}_t = (1/n)\sum_{i=1}^{n} y_{it}$$

The method is to subtract this common mean from each observation (i.e., form $y_{it}^* = y_{it} - \overline{y}_t$) and estimate (4.45) using the values of y_{it}^*. In the example at hand, y_{it} is the logarithm of real rate i at period t; hence, for each time period t, the average of these logarithmic values was subtracted from y_{it}. The last three columns of Table 4.8 show the test results for the $\{y_{it}^*\}$ sequences. Notice that the lag lengths have not changed, but the average value of the t-statistics is −2.50. As such, it is possible to reject the null hypothesis that the real rates are not stationary.

Limitations of the Panel Unit Root Test

1. The null hypothesis for the IPS test is $\gamma_i = \gamma_2 = \cdots = \gamma_n = 0$. Rejection of the null hypothesis means that *at least* one of the γ_i's differs from zero. Thus, it is possible for only one or two values of the γ_i to differ from zero and still reject the null hypothesis. Unfortunately, there is no particular way of knowing which of the γ_i are statistically different from zero. As such, the results of a panel unit root test may be dependent on the choice of the time-series variables included in the panel.
2. At this point, there is substantial disagreement about the asymptotic theory underlying the test. Sample size can approach infinity by increasing n for a given T, increasing T for a given n, or by simultaneously increasing n and T. Unfortunately, many of the important findings about the various tests are sensitive to this seemingly innocuous choice among the various assumptions. For example, the critical values reported in Table 4.7 are invariant to augmenting (4.45) with lagged changes for large T. However, for small T and large n, the critical values are dependent on the magnitudes of the various β_{ij}.
3. The test requires that the error terms from (4.45) be serially uncorrelated and contemporaneously uncorrelated. You need to determine the values of p_i to ensure that the autocorrelations of $\{\varepsilon_{it}\}$ are zero. Nevertheless, the errors may be contemporaneously correlated in that $E\varepsilon_{it}\varepsilon_{jt} \neq 0$. If the regression residuals are correlated across equations, the critical values in Table 4.7 are not applicable. The example above illustrates a common technique to correct for correlation across equations. As in the example, you can subtract a common time effect from each observation. However, there is no assurance

that this correction will completely eliminate the correlation. Moreover, it is quite possible that $\{\bar{y}_t\}$ is nonstationary. Subtracting a nonstationary component from each sequence is clearly at odds with the notion that the variables are stationary. As an alternative, many researchers would generate their own critical values by bootstrapping the value of $\bar{t}$. Some of the details regarding bootstrapping are described in Section 4.3 of the *Supplementary Manual*.

There are a number of other panel unit root tests in the literature. The Maddala–Wu (1999) test is similar to the IPS test but requires that you bootstrap your own critical values. The Levin–Lin–Chu (2002) test has the more restrictive alternative hypothesis $\gamma_1 = \gamma_2 = \cdots = \gamma_n$. Nevertheless, the cautions listed above are applicable to all of the panel unit root tests. An interesting comparison of the tests can be found in the August 2001 issue of the *Journal of Money Credit and Banking*. Three different articles perform various panel unit roots for a number of real exchange rate series.

12. TRENDS AND UNIVARIATE DECOMPOSITIONS

The findings of Nelson and Plosser (1982) suggest that many economic time series have a stochastic trend plus a stationary component. Having observed a series but not the individual components, is there any way to decompose the series into its constituent parts? Numerous economic theories suggest that it is important to distinguish between temporary and permanent movements in a series. A sale (i.e., a temporary price decline) is designed to induce us to purchase now rather than in the future. Labor economists argue that "hours supplied" is more responsive to a temporary wage increase than to a permanent increase. The idea is that workers will temporarily substitute income for leisure time. Certainly, modern theories of the consumption function that classify an individual's income into permanent and transitory components highlight the importance of such as decomposition.

Any such decomposition is straightforward if it is known that the trend in $\{y_t\}$ is purely deterministic. For example, a linear time trend induces a fixed change in each and every period. This deterministic trend can be subtracted from the actual value of y_t to obtain the stationary component.

A difficult conceptual issue arises if the trend is stochastic. For example, suppose you are asked to measure the current phase of the business cycle. If the trend in GDP is stochastic, how is it possible to tell if GDP is above or below trend? The traditional measurement of a recession by consecutive quarterly declines in real GDP is not helpful. After all, if GDP has a deterministic trend component, a negative realization for the stationary component may be outweighed by the positive deterministic trend component.

If it is possible to decompose a sequence into its separate permanent and stationary components, the issue can be solved. To better understand the nature of stochastic trends, note that—in contrast to a deterministic trend—a stochastic trend increases *on average* by a fixed amount each period. For example, consider the random walk plus drift model:

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

Since $E\varepsilon_t = 0$, the *average* change in y_t is the deterministic constant a_0. Of course, in any period t, the actual change will differ from a_0 by the stochastic quantity ε_t. Yet, each sequential change in $\{y_t\}$ adds to its level regardless of whether the change results from the deterministic or the stochastic component. As we saw in (4.5), the random walk plus drift model has no stationary component; hence, it is a model of pure trend.

The idea that a random walk plus drift is a pure trend has proved especially useful in time-series analysis. Beveridge and Nelson (1981) show how to decompose any ARIMA$(p, 1, q)$ model into the sum of a random walk plus drift and a stationary component (i.e., the general trend plus irregular model). Before considering the general case, begin with the simple example of an ARIMA(0, 1, 2) model:

$$y_t = y_{t-1} + a_0 + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} \tag{4.47}$$

If $\beta_1 = \beta_2 = 0$, (4.47) is nothing more than the pure random walk plus drift model. The introduction of the two moving average terms adds a stationary component to the $\{y_t\}$ sequence. The first step in understanding the procedure of Beveridge and Nelson (1981) is to obtain the forecast function. For now, keep the issue simple by defining $e_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2}$ so that we can write $y_t = y_{t-1} + a_0 + e_t$. Given an initial condition for y_0, the general solution for y_t is

$$y_t = a_0 t + y_0 + \sum_{i=1}^{t} e_i \tag{4.48}$$

Updating by s periods, we obtain

$$y_{t+s} = a_0(t + s) + y_0 + \sum_{i=1}^{t+s} e_i \tag{4.49}$$

Substituting (4.48) into (4.49) so as to eliminate y_0 yields

$$y_{t+s} = a_0 s + y_t + \sum_{i=1}^{s} e_{t+i} \tag{4.50}$$

To express the solution for y_{t+s} in terms of $\{\varepsilon_t\}$ rather than $\{e_t\}$, note that

$$\sum_{i=1}^{s} e_{t+i} = \sum_{i=1}^{s} \varepsilon_{t+i} + \beta_1 \sum_{i=1}^{s} \varepsilon_{t-1+i} + \beta_2 \sum_{i=1}^{s} \varepsilon_{t-2+i} \tag{4.51}$$

so that the solution for y_{t+s} can be written as

$$y_{t+s} = a_0 s + y_t + \sum_{i=1}^{s} \varepsilon_{t+1} + \beta_1 \sum_{i=1}^{s} \varepsilon_{t-1+i} + \beta_2 \sum_{i=1}^{s} \varepsilon_{t-2+i} \tag{4.52}$$

Now consider the forecast of y_{t+s} for various values of s. Since all values of $E_t\varepsilon_{t+i} = 0$ for $i > 0$, it follows that

$$\begin{aligned} E_t y_{t+1} &= a_0 + y_t + \beta_1\varepsilon_t + \beta_2\varepsilon_{t-1} \\ E_t y_{t+2} &= 2a_0 + y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1} \\ E_t y_{t+s} &= sa_0 + y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1} \end{aligned} \tag{4.53}$$

Here, the forecasts for all $s > 1$ are equal to the expression $sa_0 + y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1}$. Thus, the forecast function converges to a linear function of the forecast horizon s; the slope of the function equals a_0 and the level equals $y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1}$. This stochastic level can be called the value of the stochastic trend at t and is denoted by μ_t. This trend plus the deterministic value $a_0 s$ constitutes the forecast $E_t y_{t+s}$. There are several interesting points to note:

1. The trend is defined to be the conditional expectation of the limiting value of the forecast function. In lay terms, the trend is the "long-term" forecast. This forecast will differ at each period t as additional realizations of $\{\varepsilon_t\}$ become available. At any period t, the stationary component of the series is the difference between y_t and the trend μ_t. Hence, the stationary component of the series is

$$y_t - \mu_t = -(\beta_1 + \beta_2)\varepsilon_t - \beta_2\varepsilon_{t-1} \tag{4.54}$$

   At any point in time such that y_t is given, the trend and the stationary components are perfectly correlated (the correlation coefficient being -1).

2. By definition, ε_t is the innovation in y_t, and the variance of the innovation is σ^2. Since the change in the trend resulting from a change in ε_t is $1 + \beta_1 + \beta_2$, the variance of the innovation in the trend can exceed the variance of y_t itself. If $(1 + \beta_1 + \beta_2)^2 > 1$, the trend is more volatile than y_t since the negative correlation between μ_t and the stationary component act to smooth the $\{y_t\}$ sequence.

3. The trend is a random walk plus drift. Since the trend at t is μ_t, it follows that $\mu_t = y_t + (\beta_1 + \beta_2)\varepsilon_t + \beta_2\varepsilon_{t-1}$. Hence

$$\begin{aligned}\Delta\mu_t &= \Delta y_t + (\beta_1 + \beta_2)\Delta\varepsilon_t + \beta_2\Delta\varepsilon_{t-1} \\ &= (y_t - y_{t-1}) + (\beta_1 + \beta_2)\varepsilon_t - \beta_1\varepsilon_{t-1} - \beta_2\varepsilon_{t-2}\end{aligned}$$

   since $y_t - y_{t-1} = a_0 + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2}$,

$$\Delta\mu_t = a_0 + (1 + \beta_1 + \beta_2)\varepsilon_t$$

   Thus, $\mu_t = \mu_{t-1} + a_0 + (1 + \beta_1 + \beta_2)\varepsilon_t$, so that the trend at t is composed of the drift term a_0 plus the white-noise innovation $(1 + \beta_1 + \beta_2)\varepsilon_t$.

Beveridge and Nelson show how to recover the trend and stationary components from the data. In the example at hand, estimate the $\{y_t\}$ series using the Box–Jenkins technique. After differencing the data, an appropriately identified and estimated ARMA model will yield high-quality estimates of a_0, β_1, and β_2. Next, obtain ε_t and ε_{t-1} as the one-step-ahead forecast errors of y_t and y_{t-1}, respectively. To obtain these values, use the estimated ARMA model to make in-sample forecasts of each observation of y_{t-1} and y_t. The resulting forecast errors become ε_t and ε_{t-1}. Combining the estimated values of β_1, β_2, ε_t, and ε_{t-1} as in (4.54) yields the irregular component. Repeating for each value of t yields the entire irregular sequence. From (4.54), this irregular component is y_t minus the value of the trend; hence, the permanent component can be obtained directly.

The General ARIMA($p, 1, q$) Model

The first difference of any ARIMA($p, 1, q$) series has the stationary infinite-order moving average representation:

$$y_t - y_{t-1} = a_0 + \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots$$

As in the earlier example, it is useful to define $e_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \beta_3\varepsilon_{t-3} + \cdots$, so that it is possible to write the solution for y_{t+s} in the same form as (4.50)

$$y_{t+s} = y_t + a_0 s + \sum_{i=1}^{s} e_{t+i}$$

The next step is to express the $\{e_t\}$ sequence in terms of the various values of the $\{\varepsilon_t\}$ sequence. In this general case, (4.51) becomes

$$\sum_{i=1}^{s} e_{t+i} = \sum_{i=1}^{s} \varepsilon_{t+i} + \beta_1 \sum_{i=1}^{s} \varepsilon_{t-1+i} + \beta_2 \sum_{i=1}^{s} \varepsilon_{t-2+i} + \beta_3 \sum_{i=1}^{s} \varepsilon_{t-3+i} + \cdots \tag{4.55}$$

Since $E_t\varepsilon_{t+i} = 0$, it follows that the forecast function can be written as

$$E_t y_{t+s} = y_t + a_0 s + \left(\sum_{i=1}^{s} \beta_i\right)\varepsilon_t + \left(\sum_{i=2}^{s+1} \beta_i\right)\varepsilon_{t-1} + \left(\sum_{i=3}^{s+2} \beta_i\right)\varepsilon_{t-2} + \cdots \tag{4.56}$$

Now, to find the stochastic trend, take the limiting value of the forecast $E_t(y_{t+s} - a_0 s)$ as s becomes infinitely large. As such, the stochastic trend is

$$y_t + \left(\sum_{i=1}^{\infty} \beta_i\right)\varepsilon_t + \left(\sum_{i=2}^{\infty} \beta_i\right)\varepsilon_{t-1} + \left(\sum_{i=3}^{\infty} \beta_i\right)\varepsilon_{t-2} + \cdots$$

The key to operationalizing the decomposition is to recognize that y_{t+s} can be written as

$$y_{t+s} = \Delta y_{t+s} + \Delta y_{t+s-1} + \Delta y_{t+s-2} + \cdots + \Delta y_{t+1} + y_t$$

As such, the trend can always be written as the current value of y_t plus the sum of all of the forecasted changes in the sequence. Abstracting from $a_0 s$, the stochastic portion of the trend is

$$\begin{aligned}\lim_{s\to\infty} E_t y_{t+s} &= \lim_{s\to\infty} E_t[(y_{t+s} - y_{t+s-1}) + (y_{t+s-1} - y_{t+s-2}) + \cdots + (y_{t+2} - y_{t+1}) \\ &\quad + (y_{t+1} - y_t)] + y_t \\ &= \lim_{s\to\infty} E_t(\Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+2} + \Delta y_{t+1}) + y_t \end{aligned} \tag{4.57}$$

The useful feature of (4.57) is that the Box–Jenkins method allows you to calculate each value of $E_t\Delta y_{t+s}$. For each observation in your data set, find all s-step-ahead forecasts and construct the sum given by (4.57). Since the irregular component is y_t

minus the sum of the deterministic and stochastic trends, the irregular component can be constructed as

$$y_t - \lim_{s \to \infty} (E_t y_{t+s} + a_0 s)$$
$$= - \lim_{s \to \infty} E_t(\Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+2} + \Delta y_{t+1}) - a_0 s$$

Thus, to use the technique of Beveridge and Nelson (1981):

STEP 1: Estimate the first difference of the series using the Box–Jenkins technique. Select the best-fitting ARMA(p, q) model of the $\{\Delta y_t\}$ sequence.

STEP 2: Using the best-fitting ARMA model, for each time period $t = 1, \ldots T$, find the one-step-ahead, two-step-ahead, $\ldots$, s-step-ahead forecasts: that is, find $E_t \Delta y_{t+s}$ for each value of t and s. For each value of t, use these forecasted values to construct the sums: $E_t[\Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1}] + y_t$. In practice, it is necessary to find a reasonable approximation to (4.57); in their own work, Beveridge and Nelson let $s = 100$. For example, for the first usable observation (i.e., $t = 1$), find the sum:

$$\mu_1 = E_1(\Delta y_{101} + \Delta y_{100} + \cdots + \Delta y_2) + y_1$$

The value of y_1 plus the sum of these forecasted changes equals $E_1 y_{101}$; the stochastic portion of trend in period 1 is $E_1 y_{101} - a_0 s$ and the deterministic portion is $a_0 s$. Similarly, for $t = 2$, construct

$$\mu_2 = E_2(\Delta y_{102} + \Delta y_{101} + \cdots + \Delta y_3) + y_2$$

If there are T observations in your data set, the trend component for the last period is

$$\mu_T = E_T(\Delta y_{T+100} + \Delta y_{T+99} + \cdots + \Delta y_{T+1}) + y_T$$

The entire sequence of constructed trends (i.e., $\mu_1, \mu_2, \ldots, \mu_T$) constitutes the $\{\mu_t\}$ sequence.

STEP 3: Form the irregular component at t by subtracting the stochastic portion of the trend at t from the value of y_t. Thus, for each observation t, the irregular component is $-E_t(\Delta y_{t+100} + \Delta y_{t+99} + \cdots + \Delta y_{t+1})$.

Note that, for many series, the value of s can be quite small. For example, in the ARIMA(0, 1, 2) model of (4.47), the value of s can be set equal to 2 since all forecasts for $s > 2$ are equal to zero. If the ARMA model that is estimated in Step 1 has slowly decaying autoregressive components, the value of s should be large enough that the s-step-ahead forecasts converge to the deterministic change a_0.

Two Examples: The file PANEL.XLS contains quarterly values of the real British pound estimated by the ARIMA(0, 1, 1) process:

$$\Delta y_t = \underset{(-0.11)}{-0.0004} + \varepsilon_t + \underset{(4.75)}{0.386\varepsilon_{t-1}}$$

where Δy_t is the logarithmic change in the real British pound.

Although it is often desirable to maintain an insignificant intercept term in a regression, in this case, it is clearly undesirable since it imparts a deterministic trend into the real exchange rate. As such, reestimate the model without the intercept to obtain

$$\Delta y_t = \varepsilon_t + 0.386\varepsilon_{t-1}$$

Step 2 requires that, for each observation, we form the one-step-ahead through s-step-ahead forecasts. For this model, the mechanics are trivial since, for each period t, the one-step-ahead forecast is

$$E_t \Delta y_{t+1} = 0.386\varepsilon_t$$

and all other s-step-ahead forecasts are zero.

Thus, for each observation t, the summation $E_t(\Delta y_{t+100} + \Delta y_{t+99} + \cdots + \Delta y_{t+1})$ is equal to $0.386\varepsilon_t$. As such, for 1980Q2 (the first usable observation in the sample), the *stochastic* portion of the trend is $y_{1980Q2} + 0.386\varepsilon_{1980Q2}$ and the temporary portion of y_{1980Q2} is $-0.386\varepsilon_{1980Q2}$. Repeating for each point in the data set yields the irregular and permanent components of the sequence. The estimated ARIMA(0, 1, 1) model is the special case of (4.47) in which a_0 and β_2 are set equal to zero. As such, you should be able to write the equivalent of (4.48) through (4.54) for the real pound.

We have verified that the real U.S. GDP is the unit root process

$$\Delta lrgdp_t = 0.0078 + 0.3706\Delta lrgdp_{t-1}$$

Now, it is more difficult to calculate the sum of the forecasted changes. Nevertheless, it is worthwhile to illustrate the process for the first few values. In 1947Q2, the value of $lrgdp_t$ was 7.4776 and the value of $\Delta lrgdp_t$ was −0.00153. Since we are not interested in the deterministic portion of the trend, conditional on the information available in 1947Q2 the one-step-ahead forecast for 1947Q3 is $-5.670 \times 10^{-4} = (0.3706)(-0.00153)$ and the two-step-ahead forecast is $-2.101 \times 10^{-4} = (0.3706)(-5.670 \times 10^{-4})$. The forecasts quickly converge to zero after a few periods. Adding up all such forecasted changes, you should obtain −0.0009. Thus, abstracting from the deterministic portion of the trend, the log of real GDP is forecasted to change by −0.0009 in the very long run. Adding this sum to $lrgdp_{1947Q2}$ yields the stochastic component as $7.4476 - 0.0009 = 7.4467$. If you take the antilogs, you find the actual level of real GDP in 1947Q2 to be \$1768 billion and the permanent component to be \$1714 billion.

Repeating this process for all observations in the data set yields the time path of the trend component of real GDP. If you were to plot the trend along with the actual values, you would find that the two series virtually overlap. Since the autoregressive coefficient is so small, virtually all of the movements in the real GDP series are permanent. The cyclical component is plotted in Panel (a) of Figure 4.11. Note that the series seems to be jagged than what is normally deemed to be the business cycle. Nevertheless, the decomposed series does well in the early and late 1970s and during the financial crisis.

The Hodrick–Prescott Decomposition

Another method of decomposing a series into a trend and a stationary component has been developed by Hodrick and Prescott (1997). Suppose you observe the values y_1 through y_T and want to decompose the series into a trend $\{\mu_t\}$ and a stationary

component $y_t - \mu_t$. Consider the sum of squares

$$\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu_t)^2 + \frac{\lambda}{T}\sum_{t=2}^{T-1}[(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2$$

where λ is a constant and T is the number of usable observations.

The problem is to select the $\{\mu_t\}$ sequence so as to minimize this sum of squares. In the minimization problem, λ is an arbitrary constant reflecting the "cost" or penalty of incorporating fluctuations into the trend. In applications with quarterly data, including Hodrick and Prescott (1984) and Farmer (1993), λ is usually set equal to 1,600. Increasing the value of λ acts to "smooth out" the trend. If $\lambda = 0$, the sum of squares is minimized when $y_t = \mu_t$; the trend is equal to y_t itself. As $\lambda \to \infty$, the sum of squares is minimized when $(\mu_{t+1} - \mu_t) = (\mu_t - \mu_{t-1})$. As such, as $\lambda \to \infty$, the change in the trend is constant; the result is that there is a linear time trend. Intuitively, for large values of λ, the Hodrick–Prescott (HP) decomposition forces the change in the trend (i.e., $\Delta\mu_{t+1} - \Delta\mu_t$) to be as small as possible. This occurs when the trend is linear.

The benefit of the Hodrick–Prescott decomposition is that it uses the same method to extract the trend from a set of variables. For example, many real business cycle models indicate that all variables will have the same stochastic trend. A Beveridge–Nelson decomposition separately applied to each variable will not yield the same trend for each. Panel (b) of Figure 4.11 shows the relatively smooth cycle for the GDP series obtained from the HP filter. There is a problem in that the decomposition indicates that the economy was operating above trend in 2011 and 2012. Figure 4.12 shows the HP filter applied to real U.S. GDP, consumption, and investment. You can see that the smoothed lines (representing the trends extracted by the HP decomposition) are such that the permanent components of each series account for the majority of the variation. However, a word of warning is in order. Since the HP filter is a function that smoothes the trend, it has been shown to introduce spurious fluctuations into the

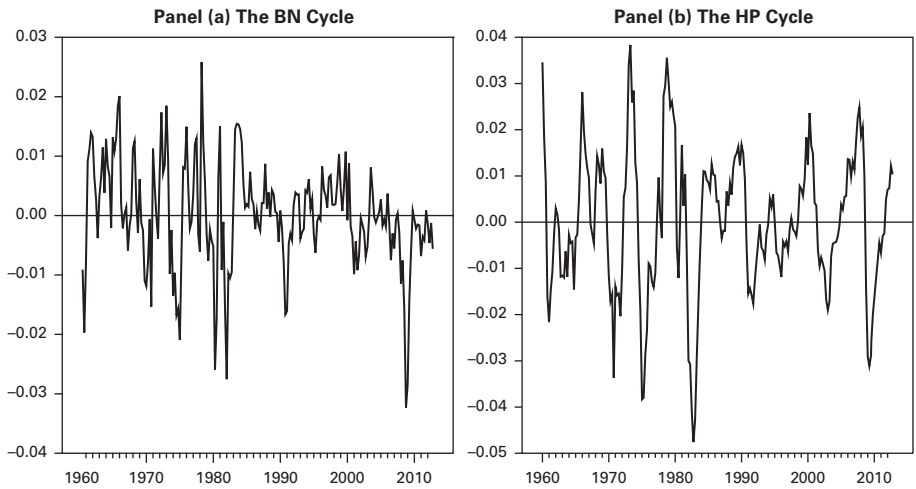


FIGURE 4.11 Two Decompositions of GDP

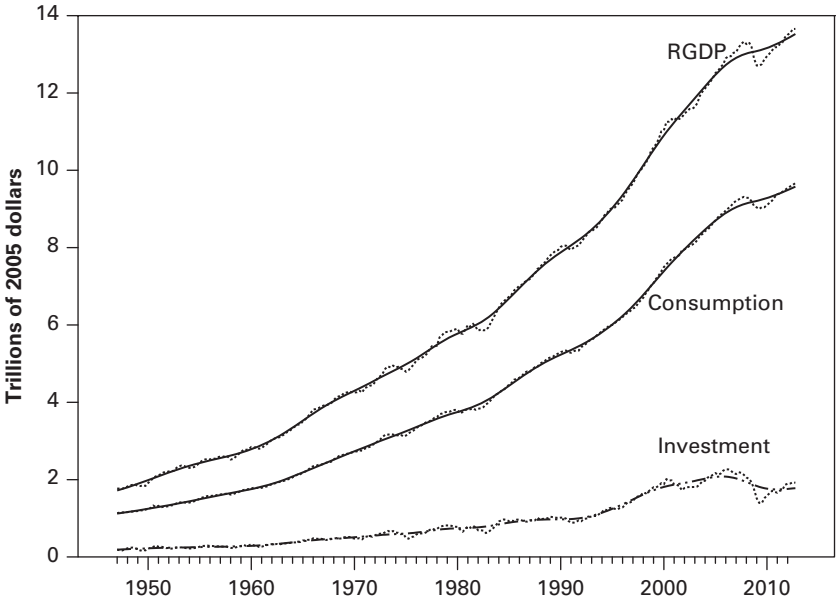


FIGURE 4.12 Real GDP, Consumption and Investment

irregular component of a series. The filter forces the stochastic trend to be a smoothed version of $(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})$. As such, the filter works best if the $\{y_t\}$ series is $I(2)$, so that smoothing the second difference of the stochastic trend is appropriate.

Note that other types of decompositions are possible. Section 4.5 of the *Supplementary Manual* examines an unobserved components decomposition of GDP into a trend and cycle.

13. SUMMARY AND CONCLUSIONS

The trend in a series can contain both stochastic and deterministic components. Differencing can remove a stochastic trend, and detrending can eliminate a deterministic trend. However, it is inappropriate to difference a trend-stationary series and to detrend a series containing a stochastic trend. The resultant irregular component of the series can be estimated using Box–Jenkins techniques.

In contrast to traditional theory, the consensus view is that most macroeconomic time series contain a stochastic trend. In finite samples, the correlogram of a unit root process will decay slowly. As such, a slowly decaying ACF can be indicative of a unit root or a near unit root process. The issue is especially important since many economic time series appear to have a nonstationary component. When you encounter such a time series, do you detrend, do you first difference, or do you do nothing since the series might be stationary?

Adherents of the Box–Jenkins methodology recommend differencing a nonstationary variable or a variable with a near unit root. For very short-term forecasts, the form of the trend is nonessential. Differencing also reveals the pattern of the other autoregressive and moving average coefficients. However, as the forecast horizon

expands, the precise form of the trend becomes increasingly important. Stationarity implies the absence of a trend and long-run mean reversion. A deterministic trend implies steady increases (or decreases) into the infinite future. Forecasts of a series with a stochastic trend converge to a steady level. As illustrated by the distinction between real business cycles and the more traditional formulations, the nature of the trend may have important theoretical implications.

The usual *t*-statistics and *F*-statistics are not applicable to determine whether or not a sequence has a unit root. Dickey and Fuller (1979, 1981) provide the appropriate test statistics to determine whether a series contains a unit root, a unit root plus drift, and/or a unit root plus drift plus a time trend. The tests can also be modified to account for seasonal unit roots. Structural breaks will bias the Dickey–Fuller test toward the nonrejection of a unit root. Perron (1989) shows how it is possible to incorporate a known structural change into the tests for unit roots. Caution needs to be exercised because it is always possible to argue that structural change has occurred; each year has something different about it than the previous year.

All the aforementioned tests have very low power to distinguish between a unit root and a near unit root process. A trend stationary process can be arbitrarily well approximated by a unit root process, and a unit root process can be arbitrarily well approximated by a trend-stationary process. Moreover, the testing procedure is confounded by the presence of the deterministic regressors (i.e., the intercept and the deterministic trend). The testing regression is misspecified if it omits any of the deterministic regressors in the data-generating process. However, too many regressors reduce the power of the tests. DF-GLS detrending methods generally have much better power than the traditional Dickey–Fuller tests. If a reasonable number of similar series are available (such as the real exchange rates from a number of countries), panel unit root tests can be used.

The fact that macroeconomic variables are not mean reverting makes it difficult to calculate the trend and cyclical components of GDP and its subcomponents. After all, traditional detrending yields nothing like a stationary cyclical component when a series contains a stochastic trend. Several methods have been devised to decompose real GDP into its permanent and temporary components. The method by Beveridge and Nelson (1981) indicates that innovations in the stochastic trend account for a sizable proportion of the period-to-period movements. However, the Beveridge–Nelson decomposition is not unique in that it forces the correlation coefficient between innovations in the trend and irregular components to have a correlation coefficient of -1. In contrast, the Hodrick–Prescott filter smoothes the trend component of a series. In Chapter 5, you will be shown a multivariate technique that allows for a unique decomposition of a series into its temporary and permanent components.

QUESTIONS AND EXERCISES

1. Given an initial condition for y_0, find the solution for y_t. Also find the s-step-ahead forecast $E_t y_{t+s}$.

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

b. $y_t = 1.1y_{t-1} + \varepsilon_t$

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

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

e. $y_t = \mu_t + \eta_t + 0.5\eta_{t-1}$, where $\mu_t = \mu_{t-1} + \varepsilon_t$
f. $y_t = \mu_t + \eta_t + 0.5\eta_{t-1}$, where $\mu_t = 0.5 + \mu_{t-1} + \varepsilon_t$
g. Can you make the models of parts b and d stationary?
h. Does model e have an ARIMA$(p, 1, q)$ representation?

2. Given the initial condition y_0, find the general solution and the forecast function (i.e., $E_t y_{t+s}$) for the following variants of the trend plus irregular model:
a. $y_t = \mu_t + v_t$, where $u_t = u_{t-1} + \varepsilon_t$, $v_t = (1 + \beta_1 L)\eta_t$, and $E\varepsilon_t\eta_t = 0$
b. $y_t = \mu_t + v_t$, where $u_t = u_{t-1} + \varepsilon_t$, $v_t = (1 + \beta_1 L)\eta_t$ and the correlation between ε_t and η_t equals unity
c. Find the ARIMA representation of each model.

3. As indicated in the text, the ACF of a series with a unit root shows little tendency to decay. Nevertheless, it may difficult to detect a unit root in a series with a negative moving average. Consider the unit root process $y_t = y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1}$.
a. Iterate backward from y_t to solve for y_t in terms of the $\{\varepsilon_t\}$ series and the initial condition y_0.
b. Use the method of undetermined coefficients to y_t in terms of the $\{\varepsilon_t\}$ series and the initial condition y_0. [Hint: The solution has the form: $y_t = \sum_{i=0}^{t-1} \alpha_i \varepsilon_{t-i} + y_0$]
c. Use your answer to part a or b to derive the first few terms of the ACF.
d. Explain how the negative MA term affects the shape of the ACF. In particular, explain how the series is "infinitely persistent" even though the coefficients of the ACF are far below unity.

4. Use the data sets that come with this text to perform the following:
a. The file PANEL.XLS contains the real exchange rates used to generate the results reported in Table 4.8. Verify the lag lengths, the values of γ, and the t-statistics reported in the left-hand side of the table.
b. Does the ERS test confirm the results you found in part a?
c. The file ERSTEST.XLS contains the data used in Section 10. Reproduce the results reported in the text.
d. The file QUARTERLY.XLS contains the M1NSA series used to illustrate the test for seasonal unit roots. Make the appropriate data transformations and verify the results concerning seasonal unit roots presented in Section 7.

5. The second column in the file BREAK.XLS contains the simulated data used in Section 8.
a. Plot the data to see if you can recognize the effects of the structural break.
b. Verify the results reported in Section 8.
c. The third column in the file BREAK.XLS contains another simulated data series called $\{y2_t\}$ with a structural break at $t = 51$. Plot the series and compare your graph to those of Figures 4.10 and 4.11.
d. Obtain the ACF and PACF of the $\{y2_t\}$ sequence and first difference of the sequence. Do the data appear to be difference stationary?
e. If you perform a Dickey–Fuller test including a constant and a trend, you should obtain

$$\begin{matrix} y2_t & = & 0.072 & - & 1.1014^{*}10^{-4}t & - & 0.022y2_{t-1} \\ & & (1.01) & & (-0.05) & & (-0.66) \end{matrix}$$

In addition to the fact that all t-statistics are small, in what other ways is this regression inadequate? What diagnostic checks would you want to perform?
f. Estimate the equation: $y2_t = a_0 + a_2 t + \mu_2 D_L$ and save the residuals. Perform a Dickey–Fuller test on the saved residuals. Perform the appropriate diagnostic tests

on this regression to ensure that the residuals approximate white noise. You should conclude that the series is a unit root process with a one-time pulse at $t = 51$.

g. Reestimate the model without the insignificant time trend. How is your answer affected?

6. The file RGDP.XLS contains the real GDP data that were used to estimate (4.29).

a. Use the series to replicate the results in Section 8.

b. It is often argued that the oil price shock in 1973 reduced the trend growth rate of real U.S. GDP. Perform the Perron test to determine whether the series is trend stationary with a break occurring in mid-1973.

c. Decompose the real GDP series into the temporary and permanent components using the HP filter and the Beveridge–Nelson decomposition. Plot the transitory component that you obtain from the HP filter and the one you obtain from the Beveridge–Nelson decomposition. In what ways are the two series different?

d. Suppose that real GDP is trend stationary with a break occurring in mid-1973. Let the deviations from trend constitute the transitory component of the series. How does this transitory component compare with your answers found in part c?

7. The file PANEL.XLS contains the real exchange rate series used to perform the panel unit root tests reported in Section 11.

a. Replicate the results of Section 11.

b. In what way do the results of the test change if Australia, France, Germany, and the United States are excluded from the panel? Why is it inappropriate to include or include countries based on their *t*-statistics?

c. Suppose that you mistakenly included a time trend in the augmented Dickey–Fuller tests. Determine how the results reported in Section 11 change.

8. The file QUARTERLY.XLS contains the U.S. interest rate data used in Section 10 of Chapter 2. Form the spread, s_t, by subtracting the *t*-bill rate from the 5-year rate. Recall that the spread appeared to be quite persistent in that $\rho_1 = 0.86$ and $\rho_2 = 0.68$.

a. One difficulty in performing a unit root test is to select the proper lag length. Using a maximum of 12 lags, estimate models of the form $\Delta s_t = a_0 + \gamma s_{t-1} + \Sigma\beta_i \Delta s_{t-i}$. Use the AIC, BIC, and general-to-specific (GTS) methods to select the appropriate lag length. You should find that the AIC, SBC, and GTS methods select lag lengths of 9, l, and 8, respectively. In this case, does the lag length matter for the Dickey–Fuller test?

b. Use a lag length of 8 and perform an augmented Dickey–Fuller test of the spread. You should find

$$\begin{matrix} \Delta s_t & = & 0.255 & - & 0.211 s_{t-1} & + & \Sigma\beta_i \Delta s_{t-i} \\ & & (3.78) & & (-4.37) & & \end{matrix}$$

Is the spread stationary?

c. Perform an augmented Dickey–Fuller test of the 5-year rate using seven lags. Is the 5-year rate stationary?

d. Perform an augmented Dickey–Fuller test of the *t*-bill rate using 11 lags. Is the *t*-bill rate stationary?

e. How is it possible that the individual rates act as $I(1)$ processes whereas the spread acts as a stationary process?

9. The file QUARTERLY.XLS contains the index of industrial production, the money supply as measured by M1, and the unemployment rate over the $1960Q1–2012Q4$ period.

a. Show that the results using this data set verify the finding of Dickey and Fuller (1981) that industrial production (INDPROD) is $I(1)$. Use the log of INDPROD and select the lag length using the general-to-specific method.

b. Perform an augmented Dickey–Fuller test on the unemployment rate (UNEMP). If you use eight lagged changes you will find

$$\begin{matrix}\Delta\text{unemp}_t &=& 0.181 &-& 0.029\text{unemp}_{t-1} &+& \Sigma\beta_i\Delta\text{unemp}_{t-i}\\ & & (2.30) & & (-2.25) & & \end{matrix}$$

Note that the t-statistic on β_8 is -2.65.

c. Now estimate the unemployment rate using only 1-lagged change. You should find

$$\begin{matrix}\Delta\text{unemp}_t &=& 0.226 &-& 0.037\text{unemp}_{t-1} &+& 0.683\Delta\text{unemp}_{t-i}\\ & & (3.36) & & (-3.43) & & (13.36)\end{matrix}$$

The residuals show only mild evidence of serial correlation. Consider

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8
0.01	−0.01	0.08	−0.10	−0.10	0.11	0.14	−0.17

What do you conclude about the stationarity of the unemployment rate?

d. Regress INDPROD on M1NSA. You should obtain

$$\begin{matrix}\text{INDPROD}_t &=& 30.48 &+& 0.04\text{M1NSA}_t\\ & & (29.90) & & (36.58)\end{matrix}$$

Examine the ACF of the residuals. Also create a scatter plot of INDPROD_t against M1NSA_t. How do you interpret the fact that $R^2 = 0.98$ and that the t-statistic on the money supply is 36.58?

10. Use the data in the file QUARTERLY.XLS to perform the following:

a. Perform the DF-GLS test using 1 lagged change of the log of INDPROD. You should find that the coefficient on γ is -2.04. (Be sure to include a time trend.)

b. Perform the DF-GLS test using eight lags of the change in UNEMP. You should find that the coefficient on γ is -1.83.

c. The SBC indicates that only one lagged change of UNEMP is appropriate. Now perform the DF-GLS test using 1-lagged change of UNEMP. In what important sense is your answer quite different from that found in part b?

11. Chapter 6 of the *Programming Manual* analyzes the real GDP data in the file QUARTERLY(2012).XLS. Unlike the real GDP data used in the text, the date in this file begin in 1960Q1. Perform parts a through e below using this shorter data set.

a. Form the log of real GDP as $ly_t = \log(\text{RGDP})$. Detrend the data with a linear time trend and form the autocorrelations.

b. Perform an augmented Dickey–Fuller test to determine whether the series is stationary. You should find that the sample value of the τ_τ statistic is -2.16. Interpret the finding that the ϕ_3-statistic is 6.34.

c. Verify the result the difference between potential and real GDP is stationary.

d. Perform the DF-GLS test on the real and the potential GDP series.

e. Compare the trends obtained from the HP filter and the Beveridge–Nelson decomposition to the values of potential GDP.

f. The *Programming Manual* applies the tests by Zivot and Andrews (1992) and Lee and Strazicich (2003) to the $ly_t = \log(\text{RGDP})$ series using data beginning in 1960Q1. Perform the test on the longer series contained in the file REAL.XLS.