21

NONSTATIONARY DATA

21.1 INTRODUCTION

Most economic variables that exhibit strong trends, such as GDP, consumption, or the price level, are not stationary and are thus not amenable to the analysis of the previous chapter. In many cases, stationarity can be achieved by simple differencing or some other simple transformation. But new statistical issues arise in analyzing nonstationary series that are understated by this superficial observation. This chapter will survey a few of the major issues in the analysis of nonstationary data.[1] We begin in Section 21.2 with results on analysis of a single nonstationary time series. Section 21.3 examines the implications of nonstationarity for analyzing regression relationship. Finally, Section 21.4 turns to the extension of the time-series results to panel data.

21.2 NONSTATIONARY PROCESSES AND UNIT ROOTS

This section will begin the analysis of nonstationary time series with some basic results for univariate time series. The fundamental results concern the characteristics of nonstationary series and statistical tests for identification of nonstationarity in observed data.

21.2.1 THE LAG AND DIFFERENCE OPERATORS

The lag operator, L, is a device that greatly simplifies the mathematics of time-series analysis. The operator defines the lagging operation,

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

From the definition,

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

It follows that

$$L^P y_t = y_{t-P},$$
$$(L^P)^Q y_t = L^{PQ} y_t = y_{t-PQ},$$
$$(L^P)(L^Q) y_t = L^P y_{t-Q} = L^{Q+P} y_t = y_{t-Q-P}.$$

[1]With panel data, this is one of the rapidly growing areas in econometrics, and the literature advances rapidly. We can only scratch the surface. Several surveys and books provide useful extensions. Three that will be very helpful are Hamilton (1994), Enders (2004), and Tsay (2005).

Finally, for the autoregressive series $y_t = \beta y_{t-1} + \varepsilon_t$, where $|\beta| < 1$, we find $(1 - \beta L)y_t = \varepsilon_t$ or

$$y_t = \left(\frac{1}{1 - \beta L}\right)\varepsilon_t = [1 + \beta L + \beta^2 L^2 + \ldots]\varepsilon_t = \sum_{s=0}^{\infty} \beta^s \varepsilon_{t-s}.$$

The first difference operator is a useful shorthand that follows from the definition of L,

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

So, for example,

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

21.2.2 INTEGRATED PROCESSES AND DIFFERENCING

A process that figures prominently in this setting is the **random walk with drift**,

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

By direct substitution,

$$y_t = \frac{\mu + \varepsilon_t}{1 - L} = \sum_{s=0}^{\infty}(\mu + \varepsilon_{t-s}).$$

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

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

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

The series y_t is said to be **integrated of order one**, denoted $I(1)$, because taking a first difference produces a stationary process. A nonstationary series is integrated of order d, denoted $I(d)$, if it becomes stationary after being first differenced d times. A generalization of the autoregressive moving average model, $y_t = \gamma y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$, would be the series

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

The resulting model is denoted an **autoregressive integrated moving-average** model, or **ARIMA** (p, d, q).[2] In full, the model would be

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

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

where

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

This result may be written compactly as

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

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

An $I(1)$ series in its raw (undifferenced) form will typically be constantly growing or wandering about with no tendency to revert to a fixed mean. Most macroeconomic flows and stocks that relate to population size, such as output or employment, are $I(1)$. An $I(2)$ series is growing at an ever-increasing rate. The price-level data in Appendix Table F5.2 and shown later appear to be $I(2)$. Series that are $I(3)$ or greater are extremely unusual, but they do exist. Among the few manifestly $I(3)$ series that could be listed, one would find, for example, the money stocks or price levels in hyperinflationary economies such as interwar Germany or Hungary after World War II.

Example 21.1 A Nonstationary Series

The nominal GDP and consumer price index variables in Appendix Table F5.2 are strongly trended, so the mean is changing over time. Figures 21.1–21.3 plot the log of the consumer price index series in Table F5.2 and its first and second differences. The original series and first differences are obviously nonstationary, but the second differencing appears to have rendered the series stationary.

The first 10 autocorrelations of the log of the GNP deflator series are shown in Table 21.1. (See Example 20.4 for details on the ACF.) The autocorrelations of the original series show the signature of a strongly trended, nonstationary series. The first difference also exhibits nonstationarity, because the autocorrelations are still very large after a lag of 10 periods. The second difference appears to be stationary, with mild negative autocorrelation

FIGURE 21.1 Quarterly Data on Log Consumer Price Index.

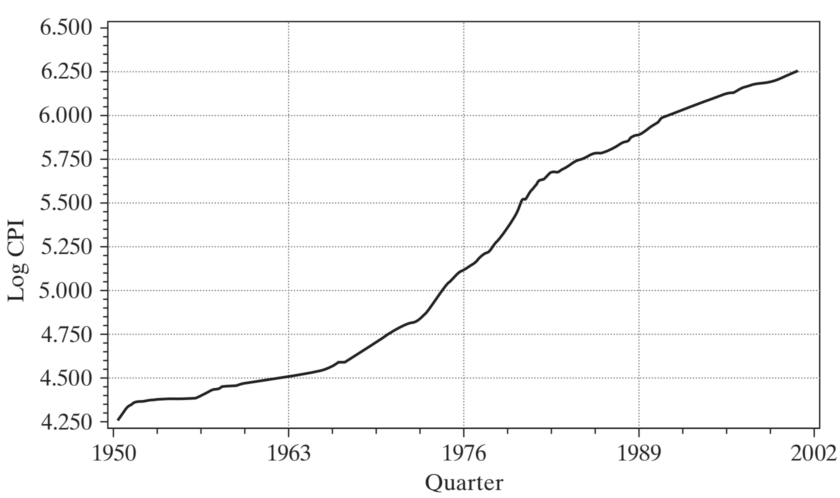
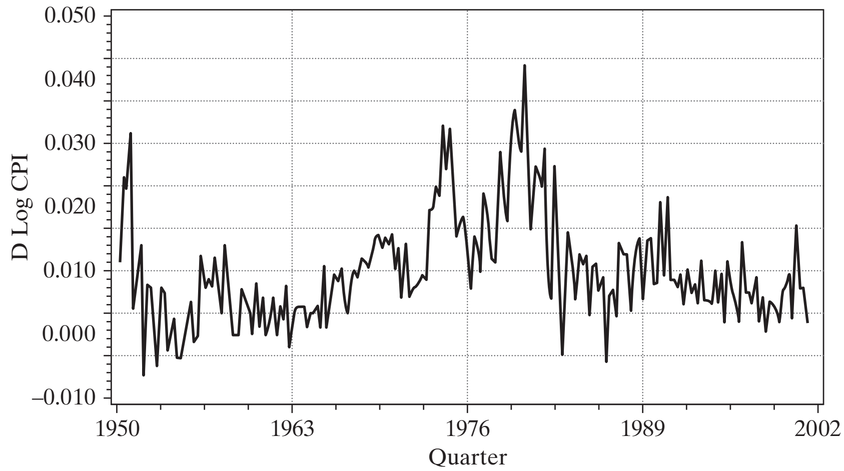


FIGURE 21.2 First Difference of Log Consumer Price Index.



at the first lag, but essentially none after that. Intuition might suggest that further differencing would reduce the autocorrelation further, but that would be incorrect. We leave as an exercise to show that, in fact, for values of γ less than about 0.5, first differencing of an AR(1) process actually increases autocorrelation.

FIGURE 21.3 Second Difference of Log Consumer Price Index.

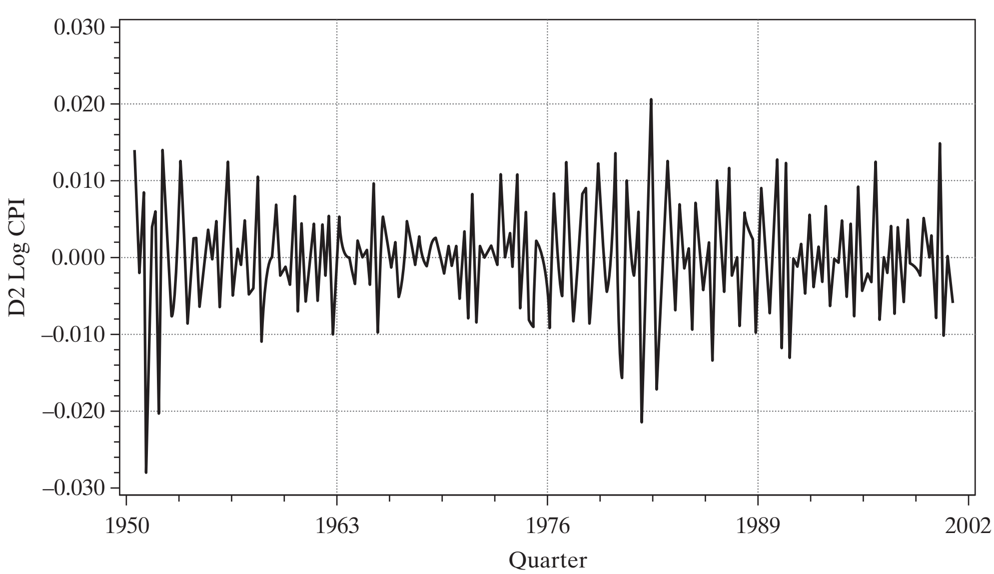


TABLE 21.1 Autocorrelations for ln Consumer Price Index

Lag	*Autocorrelation Function Original Series, log Price*		*Autocorrelation Function First Difference of log Price*		*Autocorrelation Function Second Difference of log Price*		
1	0.989	•••••••••••	0.654	•••••••	−0.422	••••••	
2	0.979	•••••••••••	0.600	•••••••	−0.111	•	
3	0.968	•••••••••••	0.621	•••••••	0.075		•
4	0.958	•••••••••••	0.600	•••••••	0.147		•
5	0.947	••••••••••	0.469	••••••	−0.112	•	
6	0.936	••••••••••	0.418	••••••	−0.037	•	
7	0.925	••••••••••	0.393	•••••	0.008		•
8	0.914	••••••••••	0.361	•••••	0.034		•
9	0.903	••••••••••	0.303	•••••	−0.023	•	
10	0.891	••••••••••	0.262	•••	−0.041	•	

21.2.3 RANDOM WALKS, TRENDS, AND SPURIOUS REGRESSIONS

In a seminal paper, Granger and Newbold (1974) argued that researchers had not paid sufficient attention to the warning of very high autocorrelation in the residuals from conventional regression models. Among their conclusions were that macroeconomic data, as a rule, were integrated and that in regressions involving the levels of such data, the standard significance tests were usually misleading. The conventional t and F tests would tend to reject the hypothesis of no relationship when, in fact, there might be none. The general result at the center of these findings is that conventional linear regression, ignoring serial correlation, of one random walk on another is virtually certain to suggest a significant relationship, even if the two are, in fact, independent. Among their extreme conclusions, Granger and Newbold suggested that researchers use a critical t value of 11.2 rather than the standard normal value of 1.96 to assess the significance of a coefficient estimate. Phillips (1986) took strong issue with this conclusion. Based on a more general model and on an analytical rather than a Monte Carlo approach, he suggested that the normalized statistic $t_\beta/\sqrt{T}$ be used for testing purposes rather than $t_{\boldsymbol{\beta}}$ itself. For the 50 observations used by Granger and Newbold, the appropriate critical value would be close to 15! If anything, Granger and Newbold were too optimistic

The random walk with drift,

$$z_t = \mu + z_{t-1} + \varepsilon_t, \qquad \textbf{(21-1)}$$

and the **trend stationary process**,

$$z_t = \mu + \beta t + \varepsilon_t, \qquad \textbf{(21-2)}$$

where, in both cases, ε_t is a white noise process, appear to be reasonable characterizations of many macroeconomic time series.[3] Clearly, both of these will produce strongly trended,

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

nonstationary series,[4] so it is not surprising that regressions involving such variables almost always produce significant relationships. The strong correlation would seem to be a consequence of the underlying trend, whether or not there really is any regression at work. But Granger and Newbold went a step further. The intuition is less clear if there is a pure **random walk** at work,

$$z_t = z_{t-1} + \varepsilon_t, \tag{21-3}$$

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

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

$$(1 - L)z_t = \alpha + v_t, \tag{21-4}$$

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

In all three settings, differencing or detrending would seem to be a natural first step. On the other hand, it is not going to be immediately obvious which is the correct way to proceed—the data are strongly trended in all three cases—and taking the incorrect approach will not necessarily improve matters. For example, first differencing in (21-1) or (21-3) produces a white noise series, but first differencing in (21-2) trades the trend for autocorrelation in the form of an MA(1) process. On the other hand, detrending—that is, computing the residuals from a regression on time—is obviously counterproductive in (21-1) and (21-3), even though the regression of z_t on a trend will appear to be significant for the reasons we have been discussing, whereas detrending in (21-2) appears to be the right approach.[5] Because none of these approaches is likely to be obviously preferable at the outset, some means of choosing is necessary. Consider nesting all three models in a single equation,

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

Now subtract z_{t-1} from both sides of the equation and introduce the artificial parameter γ, Then,

$$\begin{aligned} z_t - z_{t-1} &= \mu\gamma + \beta\gamma t + (\gamma - 1)z_{t-1} + \varepsilon_t \\ &= \alpha_0 + \alpha_1 t + (\gamma - 1)z_{t-1} + \varepsilon_t, \end{aligned} \tag{21-5}$$

where, by hypothesis, $\gamma = 1$. Equation (21-5) provides the basis for a variety of tests for unit roots in economic data. In principle, a test of the hypothesis that $\gamma - 1$ equals zero gives confirmation of the random walk with drift, because if γ equals 1 (and α_1 equals zero), then (21-1) results. If $\gamma - 1$ is less than zero, then the evidence favors the trend stationary (or some other) model, and detrending (or some alternative) is the preferable

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

[5]See Nelson and Kang (1984).

approach. The practical difficulty is that standard inference procedures based on least squares and the familiar test statistics are not valid in this setting. The issue is discussed in the next section.

21.2.4 TESTS FOR UNIT ROOTS IN ECONOMIC DATA

The implications of unit roots in macroeconomic data are, at least potentially, profound. If a structural variable, such as real output, is truly $I(1)$, then shocks to it will have permanent effects. If confirmed, then this observation would mandate some rather serious reconsideration of the analysis of macroeconomic policy. For example, the argument that a change in monetary policy could have a transitory effect on real output would vanish.[6] The literature is not without its skeptics, however. This result rests on a razor's edge. Although the literature is thick with tests that have failed to reject the hypothesis that $\gamma = 1$, many have also not rejected the hypothesis that $\gamma \geq 0.95$, and at 0.95 (or even at 0.99), the entire issue becomes moot.[7]

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

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

The downward bias of the least squares estimator when γ approaches one has been widely documented.[8] For $|\gamma| < 1$, however, the least squares estimator,

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

does have

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

and

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

Does the result hold up if $\gamma = 1$? The case is called the unit root case, because in the ARMA representation $C(L)y_t = \varepsilon_t$, the characteristic equation $1 - \gamma z = 0$ has one root equal to one. That the limiting variance appears to go to zero should raise suspicions. The literature on the question dates back to Mann and Wald (1943) and Rubin (1950). But for econometric purposes, the literature has a focal point at the celebrated papers of Dickey and Fuller (1979, 1981). They showed that if γ equals one, then

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

where v is a random variable with finite, positive variance, and in finite samples, $E[c] < 1$.[9]

There are two important implications in the Dickey–Fuller results. First, the estimator of γ is biased downward if γ equals one. Second, the OLS estimator of γ converges to its

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

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

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

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

probability limit more rapidly than the estimators to which we are accustomed. That is, the variance of c under the null hypothesis is $O(1/T^2)$, not $O(1/T)$. (In a mean squared error sense, the OLS estimator is superconsistent.) It turns out that the implications of this finding for the regressions with trended data are considerable.

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

21.2.5 THE DICKEY–FULLER TESTS

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

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

Under the null hypothesis that $\gamma = 1$, there are two approaches to carrying out the test. The conventional t ratio, $\text{DF}_\tau = (\hat{\gamma} - 1)/\text{Est. Std. Error}(\hat{\gamma})$, with the revised set of critical values may be used for a one-sided test. Critical values for this test are shown in the top panel of Table 21.2. Note that, in general, the critical value is considerably larger in absolute value than its counterpart from the t distribution. The second approach is based on the statistic $\text{DF}_\gamma = T(\hat{\gamma} - 1)$. Critical values for this test are shown in the top panel of Table 21.3.

The simple random walk model is inadequate for many series. Consider the rate of inflation from 1950.2 to 2000.4 (plotted in Figure 21.4) and the log of GDP over the same period (plotted in Figure 21.5). The first of these may be a random walk, but it is clearly drifting. The log GDP series, in contrast, has a strong trend. For the first of these, a random walk with drift may be specified,

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

or

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

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

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

or

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

TABLE 21.2 Critical Values for the Dickey–Fuller DFτ Test.

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

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

The tests for these forms may be carried out in the same fashion. For the model with drift only, the center panels of Tables 21.2 and 21.3 are used. When the trend is included, the lower panel of each table is used.

Example 21.2 Tests for Unit Roots

Cecchetti and Rich (2001) studied the effect of monetary policy on the U.S. economy. The data used in their study were the following variables:

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

Data used in their analysis were from the period 1959.1 to 1997.4. As part of their analysis, they checked each of these series for a unit root and suggested that the hypothesis of a unit root could only be rejected for the last two variables. We will reexamine these data for the longer interval, 1950II to 2000IV. The data are in Appendix Table F5.2. Figures 21.6–21.9 show

TABLE 21.3 Critical Values for the Dickey–Fuller DFτ Test.

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

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

FIGURE 21.4 Rate of Inflation in the Consumer Price Index.

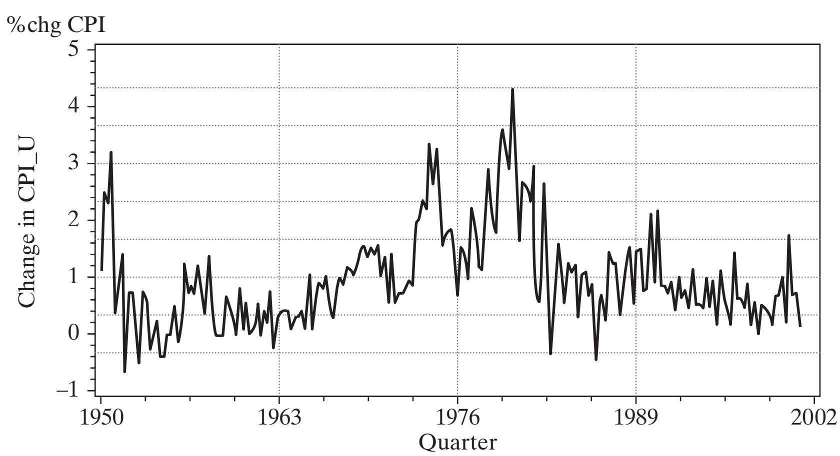
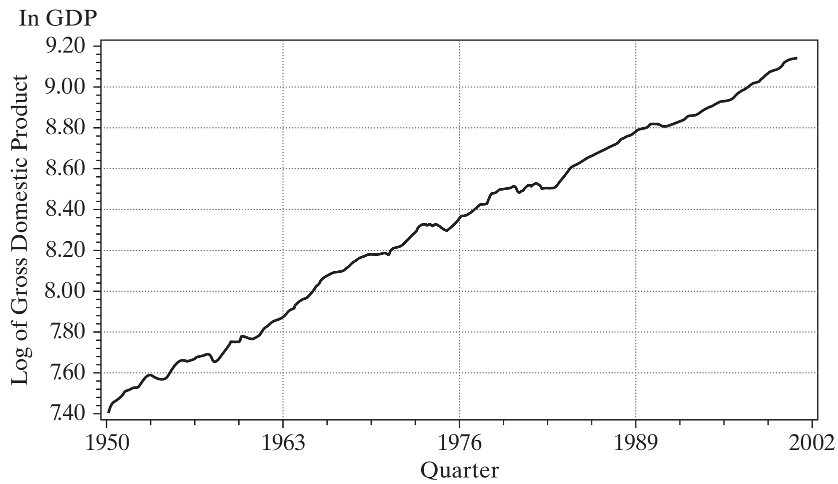


FIGURE 21.5 Log of Gross Domestic Product.



the behavior of the last four variables. The first two are shown in Figures 21.4 and 21.5. Only the real output figure shows a strong trend, so we will use the random walk with drift for all the variables except this one.

The Dickey–Fuller tests are carried out in Table 21.4. There are 203 observations used in each one. The first observation is lost when computing the rate of inflation and the change in the money stock, and one more is lost for the difference term in the regression. The critical values from interpolating to the second row, last column in each panel for 95% significance

FIGURE 21.6 T-Bill Rate.

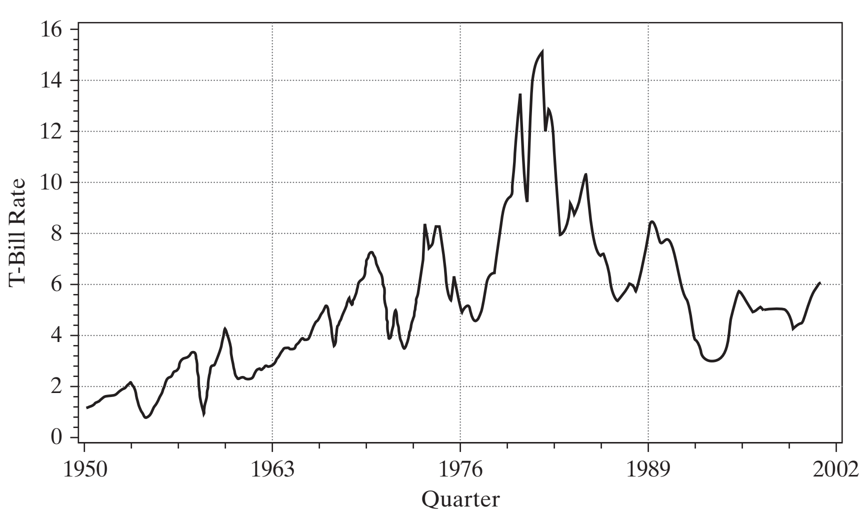
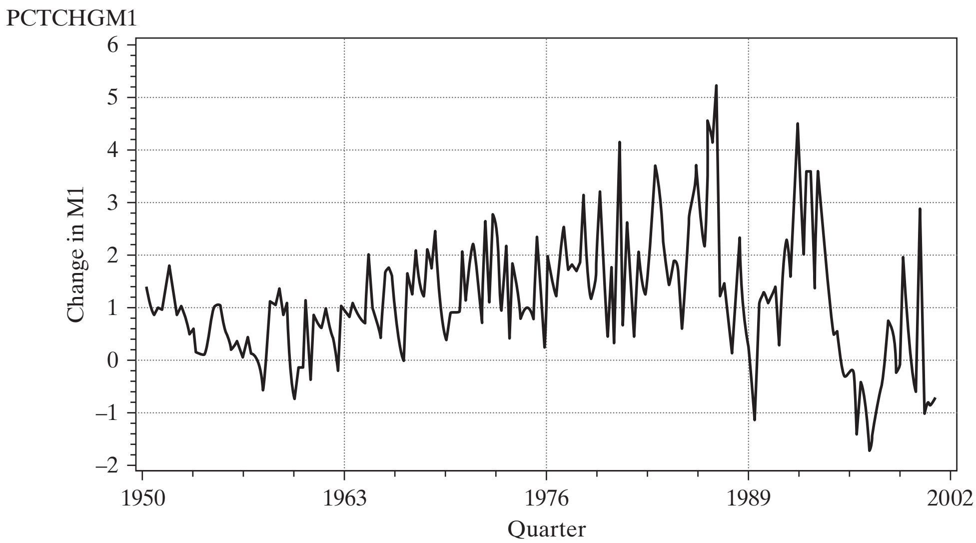


FIGURE 21.7 Percentage Change in the Money Stock.



and a one-tailed test are −3.68 and −24.2, respectively, for DF_τ and DF_γ for the output equation, which contains the time trend, and −3.14 and −16.8 for the other equations, which contain a constant but no trend. For the output equation (y), the test statistics are

$$DF_\tau = \frac{0.9584940384 - 1}{.017880922} = -2.32 > -3.44,$$

FIGURE 21.8 Ex-Post Real T-Bill Rate.

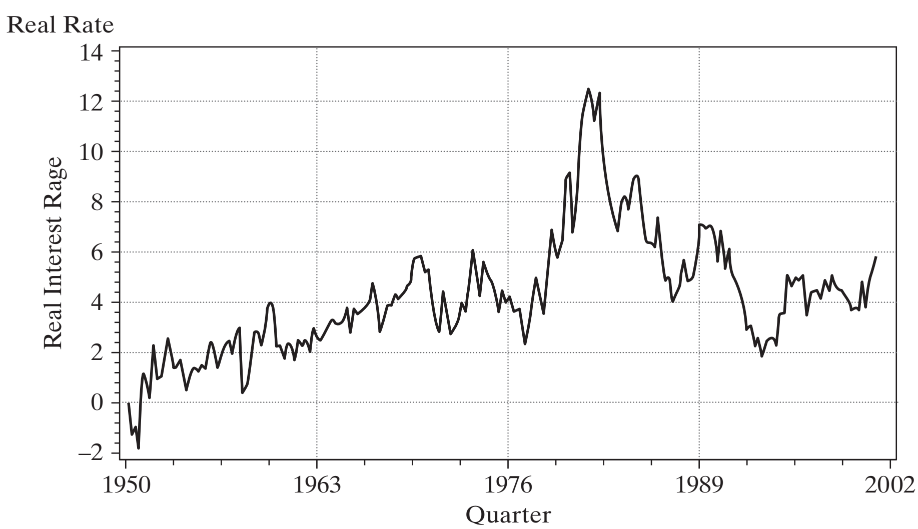
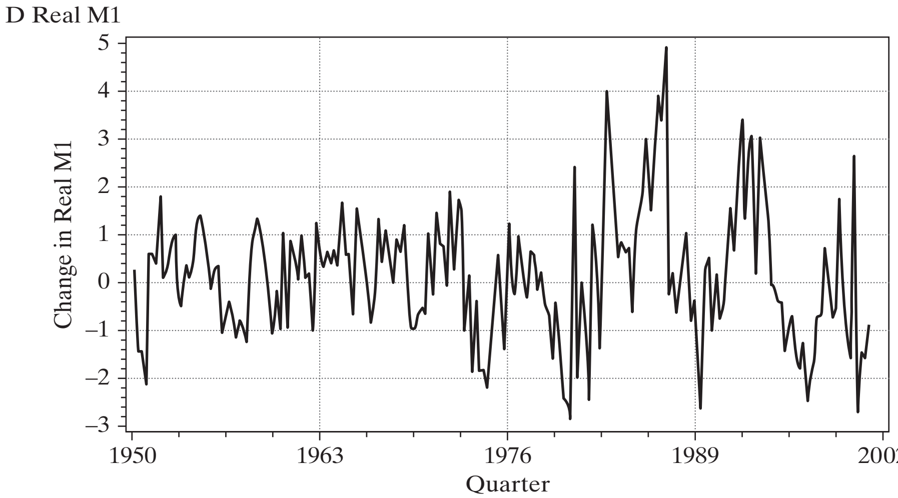


FIGURE 21.9 Change in the Real Money Stock.



and

$$DF_\gamma = 202(0.9584940384 - 1) = -8.38 > -21.2.$$

Neither is less than the critical value, so we conclude (as have others) that there is a unit root in the log GDP process. The results of the other tests are shown in Table 21.4. Surprisingly, these results do differ sharply from those obtained by Cecchetti and Rich (2001) for π and Δm. The sample period appears to matter; if we repeat the computation using Cecchetti and Rich's interval, 1959.4 to 1997.4, then DF_τ equals -3.51. This is borderline, but less contradictory. For Δm, we obtain a value of -4.204 for DF_τ when the sample is restricted to the shorter interval.

TABLE 21.4 Unit Root Tests (standard errors of estimates in parentheses).

	μ	β	γ	DF_τ	DF_γ	*Conclusion*
π	0.332		0.659	−6.40	−68.88	Reject H_0
	(0.0696)		(0.0532)	$R^2 = 0.432$	$s = 0.643$	
y	0.320	0.00033	0.958	−2.35	−8.48	Do not reject H_0
	(0.134)	(0.00015)	(0.0179)	$R^2 = 0.999$	$s = 0.001$	
i	0.228		0.961	−2.14	−7.88	Do not reject H_0
	(0.109)		(0.0182)	$R^2 = 0.933$	$s = 0.743$	
Δm	0.448		0.596	−7.05	−81.61	Reject H_0
	(0.0923)		(0.0573)	$R^2 = 0.351$	$s = 0.929$	
$i - \pi$	0.615		0.557	−7.57	−89.49	Reject H_0
	(0.185)		(0.0585)	$R^2 = 0.311$	$s = 2/395$	
$\Delta m - \pi$	0.0700		0.490	−8.25t	−103.02	Reject H_0
	(0.0833)		(0.0618)	$R^2 = 0.239$	$s = 1.176$	

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

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

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

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

exactly as constructed before, and

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

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

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

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

where

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

The unit root test is carried out as before by testing the null hypothesis $\gamma^* = 0$ against $\gamma^* < 0$.[10] The t test, DF_τ, may be used. If the failure to reject the unit root is taken as evidence that a unit root is present, that is, $\gamma^* = 0$, then the model specializes to the AR($p - 1$) model in the first differences, which is an ARIMA($p - 1, 1, 0$) model for y_t. For a model with a time trend,

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

the test is carried out by testing the joint hypothesis that $\beta = \gamma^* = 0$. Dickey and Fuller (1981) present counterparts to the critical F statistics for testing the hypothesis. Some of their values are reproduced in the first row of Table 21.2. (Authors frequently focus on γ^* and ignore the time trend, maintaining it only as part of the appropriate formulation. In this case, one may use the simple test of $\gamma^* = 0$ as before, with the DF_τ critical values.)

The lag length, p, remains to be determined. As usual, we are well advised to test down to the right value instead of up. One can take the familiar approach and sequentially examine the t statistic on the last coefficient—the usual t test is appropriate.

[10]It is easily verified that one of the roots of the characteristic polynomial is $1/(\gamma_1 + \gamma_2 + \ldots + \gamma_P)$.

An alternative is to combine a measure of model fit, such as the regression s^2, with one of the information criteria. The Akaike and Schwarz (Bayesian) information criteria would produce the two information measures

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

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

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

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

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

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

gave good results.

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

$$y_t = \delta_t + \gamma y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_p \Delta y_{t-p} + \varepsilon_t, \tag{21-6}$$

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

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

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

where

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

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

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

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

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

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

Elliot, Rothenberg, and Stock (1996) have proposed a method they denote the ADF-GLS procedure, which is designed to accommodate more general formulations of ε; the process generating ε_t is assumed to be an $I(0)$ stationary process, possibly an ARMA(r, s). The null hypothesis, as before, is $\gamma = 1$ in (21-6) where $\delta_t = \mu$ or $\mu + \beta t$. The method proceeds as follows:

Step 1. Linearly regress

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

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

Step 2. The Dickey–Fuller DF_τ test can now be carried out using the model

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

If the model does not contain the time trend, then the t statistic for $(\gamma - 1)$ may be referred to the critical values in the center panel of Table 21.2. For the trend stationary model, the critical values are given in a table presented in Elliot et al. The 97.5% critical values for a one-tailed test from their table is -3.15.

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

Example 21.3 Augmented Dickey–Fuller Test for a Unit Root in GDP

Dickey and Fuller (1981) apply their methodology to a model for the log of a quarterly series on output, the Federal Reserve Board Production Index. The model used is

$$y_t = \mu + \beta t + \gamma y_{t-1} + \phi(y_{t-1} - y_{t-2}) + \varepsilon_t. \qquad \textbf{(21-7)}$$

The test is carried out by testing the joint hypothesis that both β and γ^* are zero in the model

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

(If $\gamma = 0$, then μ^* will also by construction.) We will repeat the study with our data on real GDP from Appendix Table F5.2 using observations 1950.1–2000.4.

We will use the augmented Dickey–Fuller test first. Thus, the first step is to determine the appropriate lag length for the augmented regression. Using Schwert's suggestion, we find that the maximum lag length should be allowed to reach $p_{max} =$ [integer part of $12(204/100)^{0.25}$] $= 14$.

The specification search uses observations 18 to 204, because as many as 17 coefficients will be estimated in the equation

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

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

The linear regression results for the equation in (21-7) are

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

The two test statistics are

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

and

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

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

For the Phillips and Perron statistic, we need several additional intermediate statistics. Following Hamilton (1994, p. 512), we choose $L = 4$ for the long-run variance calculation. Other values we need are $T = 202$, $\hat{\gamma} = 0.9516613$, $s^2 = 0.00008311488$, $v^2 = 0.00027942647$, and the first five autocovariances, $c_0 = 0.000081469$, $c_1 = -0.00000351162$, $c_2 = 0.00000688053$, $c_3 = 0.000000597305$, and $c_4 = -0.00000128163$. Applying these to the weighted sum produces $a = 0.0000840722$, which is only a minor correction to c_0. Collecting the results, we obtain the Phillips–Perron statistics, $Z_\tau = -2.89921$ and $Z_\gamma = -15.44133$. Because these are applied to the same critical values in the Dickey–Fuller tables, we reach the same conclusion as before—we do not reject the hypothesis of a unit root in log GDP.

21.2.6 THE KPSS TEST OF STATIONARITY

Kwiatkowski et al. (1992) (KPSS) have devised an alternative to the Dickey–Fuller test for stationarity of a time series. The procedure is a test of nonstationarity against the null hypothesis of stationarity in the model

$$\begin{aligned} y_t &= \alpha + \beta t + \gamma \sum_{i=1}^{t} z_i + \varepsilon_t, \qquad t = 1, \ldots, T \\ &= \alpha + \beta t + \gamma Z_t + \varepsilon_t, \end{aligned}$$

where ε_t is a stationary series and z_t is an i.i.d. stationary series with mean zero and variance one. (These are merely convenient normalizations because a nonzero mean would move to α and a nonunit variance is absorbed in γ.) If γ equals zero, then the process is stationary if $\beta = 0$ and trend stationary if $\beta \neq 0$. Because Z_t, is $I(1)$, y_t is nonstationary if γ is nonzero.

The KPSS test of the null hypothesis, $H_0: \gamma = 0$, against the alternative that γ is nonzero reverses the strategy of the Dickey–Fuller statistic (which tests the null hypothesis $\gamma < 1$ against the alternative $\gamma = 1$). Under the null hypothesis, α and β can be estimated by OLS. Let e_t denote the tth OLS residual,

$$e_t = y_t - a - bt,$$

and let the sequence of partial sums be

$$E_t = \sum_{s=1}^{t} e_s, \quad t = 1, \ldots, T.$$

(Note $E_T = 0$.) The KPSS statistic is

$$\text{KPSS} = \frac{\sum_{t=1}^{T} E_t^2}{T^2 \hat{\sigma}^2},$$

where

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} e_t^2}{T} + 2\sum_{j=1}^{L}\left(1 - \frac{j}{L+1}\right) r_j,$$

$$r_j = \frac{\sum_{s=j+1}^{T} e_s e_{s-j}}{T},$$

and L is chosen by the analyst. [See (20-17).] Under normality of the disturbances, ε_t, the KPSS statistic is an LM statistic. The authors derive the statistic under more general conditions. Critical values for the test statistic are estimated by simulation. The 0.05 upper-tail values reported by the authors (in their Table 1, p. 166) for $\beta = 0$ and $\beta \neq 0$ are 0.463 and 0.146, respectively.

Example 21.4 Is There a Unit Root in GDP?

Using the data used for the Dickey–Fuller tests in Example 21.3, we repeated the procedure using the KPSS test with $L = 10$. The two statistics are 1.953 without the trend and 0.312 with it. Comparing these results to the values in Table 21.4 we conclude (again) that there is, indeed, a unit root in ln GDP. Or, more precisely, we conclude that ln GDP is not a stationary series, nor even a trend stationary series.

21.3 COINTEGRATION

Studies in empirical macroeconomics almost always involve nonstationary and trending variables, such as income, consumption, money demand, the price level, trade flows, and exchange rates. Accumulated wisdom and the results of the previous sections suggest that the appropriate way to manipulate such series is to use differencing and other transformations (such as seasonal adjustment) to reduce them to stationarity and then to analyze the resulting series as VARs or with the methods of Box and Jenkins (1984). But recent research and a growing literature have shown that there are more interesting, appropriate ways to analyze trending variables.

In the *fully specified* regression model,

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

there is a presumption that the disturbances ε_t are a stationary, white noise series.[11] But this presumption is unlikely to be true if y_t and x_t are integrated series. Generally, if two series are integrated to different orders, then linear combinations of them will be integrated to the higher of the two orders. Thus, if y_t and x_t are $I(1)$—that is, if both are trending variables—then we would normally expect $y_t - \beta x_t$ to be $I(1)$ regardless of the value of β, not $I(0)$ (i.e., not stationary). If y_t and x_t are each drifting upward with their own trend, then unless there is some relationship between those trends, the difference between them should also be growing, with yet another trend. There must be some kind of inconsistency in the model. On the other hand, if the two series are both $I(1)$, then there *may* be a β such that

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

is $I(0)$. Intuitively, if the two series are both $I(1)$, then this partial difference between them might be stable around a fixed mean. The implication would be that the series are drifting together at roughly the same rate. Two series that satisfy this requirement are said to be **cointegrated**, and the vector $[1, -\beta]$ (or any multiple of it) is a *cointegrating vector*. In such a case, we can distinguish between a long-run relationship between y_t and x_t, that is, the manner in which the two variables drift upward together, and the short-run dynamics, that is, the relationship between deviations of y_t from its long-run trend and deviations of x_t from its long-run trend. If this is the case, then differencing of the data would be counterproductive, because it would obscure the long-run relationship between y_t and x_t. Studies of cointegration and a related technique, error correction, are concerned with methods of estimation that preserve the information about both forms of covariation.[12]

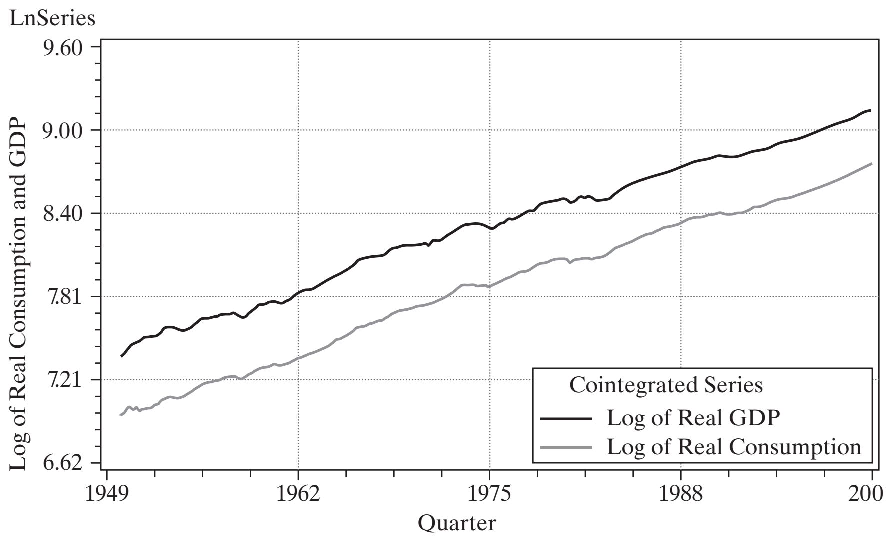
Example 21.5 Cointegration in Consumption and Output

Consumption and income provide one of the more familiar examples of the phenomenon described previously. The logs of GDP and consumption for 1950.1 to 2000.4 are plotted in Figure 21.10. Both variables are obviously nonstationary. We have already verified that there is a unit root in the income data. We leave as an exercise for the reader to verify that the consumption variable is likewise $I(1)$. Nonetheless, there is a clear relationship between consumption and output. Consider a simple regression of the log of consumption on the log of income, where both variables are manipulated in mean deviation form (so, the regression includes a constant). The slope in that regression is 1.056765. The residuals from the regression, $u_t = [\ln \text{Cons}^*, \ln \text{GDP}^*][1, -1.056765]'$ (where the "*" indicates mean deviations) are plotted in Figure 21.11. The trend is clearly absent from the residuals. But it remains to verify whether the series of residuals is stationary. In the ADF regression of the least squares residuals on a constant (random walk with drift), the lagged value and the lagged first difference, the coefficient on u_{t-1} is 0.838488 (0.0370205) and that on $u_{t-1} - u_{t-2}$ is -0.098522. (The constant differs trivially from zero because two observations are lost in computing the ADF regression.) With 202 observations, we find $\text{DF}_\tau = -4.63$ and $\text{DF}_\gamma = -29.55$. Both are well below the critical values, which suggests that the residual series does not contain a unit root. We conclude (at least it appears so) that even after

[11] Any autocorrelation in the model has been removed through an appropriate transformation.

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

FIGURE 21.10 Cointegrated Variables: Logs of Consumption and GDP.

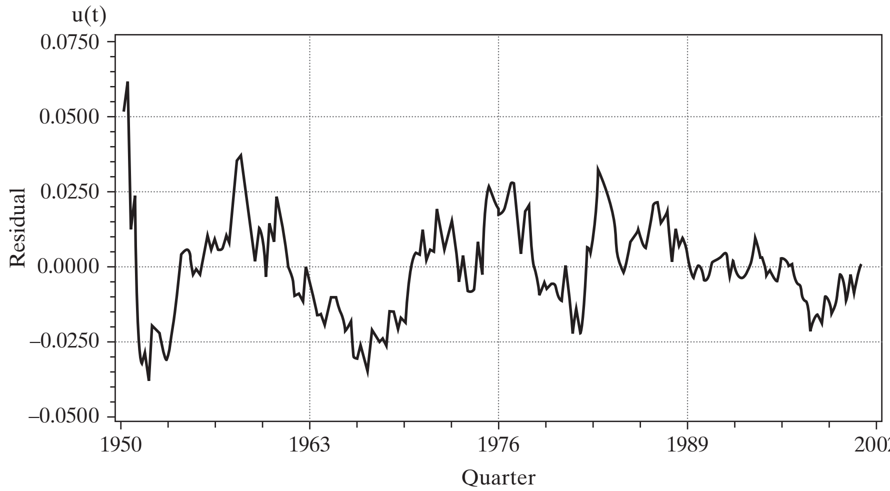


accounting for the trend, although neither of the original variables is stationary, there is a linear combination of them that is. If this conclusion holds up after a more formal treatment of the testing procedure, we will conclude that log GDP and log consumption are cointegrated.

Example 21.6 Several Cointegrated Series

The theory of purchasing power parity specifies that in long-run equilibrium, exchange rates will adjust to erase differences in purchasing power across different economies. Thus, if

FIGURE 21.11 Residuals from Consumption—Income Regression.



p_1 and p_0 are the price levels in two countries and E is the exchange rate between the two currencies, then in equilibrium,

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

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

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

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

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

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

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

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

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

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

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

Example 21.7 Multiple Cointegrating Vectors

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

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

The price level and real income are assumed to be $I(1)$. The existence of long-run equilibrium in the money market implies a cointegrating vector $\boldsymbol{\alpha}_1$. If the Fed follows a certain feedback rule, increasing the money stock when *nominal* income $(y + p)$ is low and decreasing it when nominal income is high—which might make more sense in terms of rates of growth—then there is a second cointegrating vector in which $\gamma_1 = \gamma_2$ and $\gamma_3 = 0$. Suppose that we label this vector $\boldsymbol{\alpha}_2$. The parameters in the money demand equation, notably the interest elasticity, are interesting quantities, and we might seek to estimate $\boldsymbol{\alpha}_1$ to learn the value of this quantity. Because every linear combination of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ is a cointegrating vector, to this point we are only able to estimate a hash of the two cointegrating vectors.

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

$$\boldsymbol{\alpha}_1 = [1, -\gamma_{10}, -\gamma_{11}, -\gamma_{12}, -\gamma_{13}]'$$

and

$$\boldsymbol{\alpha}_2 = [1, -\gamma_{20}, \gamma_{21}, \gamma_{21}, 0]'.$$

Although it is true that every linear combination of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ is a cointegrating vector, only the original two vectors, as they are, have a 1 in the first position of both and a 0 in the last position of the second. (The equality restriction actually overidentifies the parameter matrix.) This result is, of course, exactly the sort of analysis that we used in establishing the identifiability of a simultaneous equations system in Chapter 10.

21.3.1 COMMON TRENDS

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

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

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

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

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

w_{it}, $i = 1, 2$. Any linear combination of y_{1t} and y_{2t} must involve *both* random walks. It is clear that they cannot be cointegrated unless, in fact, $w_{1t} = w_{2t}$. That is, once again, they must have a **common trend**. Finally, suppose that y_{1t} and y_{2t} share two common trends,

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

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

As Stock and Watson determined, the preceding is the crux of the cointegration of economic variables. A set of M variables that are cointegrated can be written as a stationary component plus linear combinations of a smaller set of common trends. If the cointegrating rank of the system is r, then there can be up to $M - r$ linear trends and $M - r$ common random walks.[13] (The two-variable case is special. In a two-variable system, there can be only one common trend in total.) The effect of the cointegration is to purge these common trends from the resultant variables.

21.3.2 ERROR CORRECTION AND VAR REPRESENTATIONS

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

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

describes the variation in y_t around its long-run trend in terms of a set of $I(0)$ exogenous factors $\mathbf{x}_t$, the variation of z_t around its long-run trend, and the error correction $(y_t - \theta z_t)$, which is the equilibrium error in the model of cointegration. There is a tight connection between models of cointegration and models of error correction. The model in this form is reasonable as it stands, but in fact, it is only internally consistent if the two variables are cointegrated. If not, then the third term, and hence the right-hand side, cannot be $I(0)$, even though the left-hand side must be. The upshot is that the same assumption that we make to produce the cointegration implies (and is implied by) the existence of an error correction model.[14] As we will examine in the next section, the utility of this representation is that it suggests a way to build an elaborate model of the long-run variation in y_t as well as a test for cointegration. Looking ahead, the preceding suggests that residuals from an estimated cointegration model—that is, estimated equilibrium errors—can be included in an elaborate model of the long-run covariation of y_t and z_t. Once again, we have the foundation of Engel and Granger's approach to analyzing cointegration.

Pesaran, Shin, and Smith (2001) suggest a method of testing for a relationship in levels between a y_t and an $\mathbf{x}_t$ when there exist significant lags in the error correction form. Their **bounds test** accommodates the possibility that the regressors may be trend or difference stationary. The critical values they provide give a band that covers the polar cases in which all regressors are $I(0)$, or are $I(1)$, or are mutually cointegrated. The

[13]See Hamilton (1994, p. 578).

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

statistic is able to test for the existence of a levels equation regardless of whether the variables are $I(0)$, $I(1)$, or are cointegrated. In their application, y_t is real earnings in the UK while $\mathbf{x}_t$ includes a measure of productivity, the unemployment rate, unionization of the workforce, a *replacement ratio* that measures the difference between unemployment benefits and real wages, and a *wedge* between the real product wage and the real consumption wage. It is found that wages and productivity have unit roots. The issue then is to discern whether unionization, the wedge, and the unemployment rate, which might be $I(0)$, have level effects in the model.

Consider the vector autoregression, or VAR representation of the model

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix},$$

or

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

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

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

or

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

If all variables are $I(1)$, then all M variables on the left-hand side are $I(0)$. Whether those on the right-hand side are $I(0)$ remains to be seen. The matrix $\mathbf{\Pi}$ produces linear combinations of the variables in $\mathbf{y}_t$. But as we have seen, not all linear combinations can be cointegrated. The number of such independent linear combinations is $r < M$. Therefore, although there must be a VAR representation of the model, cointegration implies a restriction on the rank of $\mathbf{\Pi}$. It cannot have full rank; its rank is r. From another viewpoint, a different approach to discerning cointegration is suggested. Suppose that we estimate this model as an unrestricted VAR. The resultant coefficient matrix should be short-ranked. The implication is that if we fit the VAR model and impose short rank on the coefficient matrix as a restriction—how we could do that remains to be seen—then if the variables really are cointegrated, this restriction should not lead to a loss of fit. This implication is the basis of Johansen's (1988) and Stock and Watson's (1988) analysis of cointegration.

21.3.3 TESTING FOR COINTEGRATION

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

Let $\mathbf{y}_t$ denote the set of M variables that are believed to be cointegrated. Step one of either analysis is to establish that the variables are indeed integrated to the same order.

The Dickey–Fuller tests discussed in Section 21.2.4 can be used for this purpose. If the evidence suggests that the variables are integrated to different orders or not at all, then the specification of the model should be reconsidered.

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

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

which we may rewrite as

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

We can obtain estimates of $\boldsymbol{\theta}_i$ by least squares regression. If the theory is correct *and* if this OLS estimator is consistent, then residuals from this regression should estimate the equilibrium errors. There are two obstacles to consistency. First, because both sides of the equation contain $I(1)$ variables, the problem of spurious regressions appears. Second, a moment's thought should suggest that what we have done is extract an equation from an otherwise ordinary simultaneous equations model and propose to estimate its parameters by ordinary least squares. As we examined in Chapter 10, consistency is unlikely in that case. It is one of the extraordinary results of this body of theory that in this setting, neither of these considerations is a problem. In fact, as shown by a number of authors,[15] not only is $\mathbf{c}_i$, the OLS estimator of $\boldsymbol{\theta}_i$, consistent, it is **superconsistent** in that its asymptotic variance is $O(1/T^2)$ rather than $O(1/T)$ as in the usual case. Consequently, the problem of spurious regressions disappears as well. Therefore, the next step is to estimate the cointegrating vector(s), by OLS. Under all the assumptions thus far, the residuals from these regressions, e_{it}, are estimates of the equilibrium errors, ε_{it}. As such, they should be $I(0)$. The natural approach would be to apply the familiar Dickey–Fuller tests to these residuals. The logic is sound, but the Dickey–Fuller tables are inappropriate for these estimated errors. Estimates of the appropriate critical values for the tests are given by Engle and Granger (1987), Engle and Yoo (1987), Phillips and Ouliaris (1990), and Davidson and MacKinnon (1993). If autocorrelation in the equilibrium errors is suspected, then an augmented Engle and Granger test can be based on the template

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

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

Example 21.8 Cointegration in Consumption and Output

In the example presented at the beginning of this discussion, we proposed precisely the sort of test suggested by Phillips and Ouliaris (1990) to determine if (log) consumption and (log)

[15] See, for example, Davidson and MacKinnon (1993).

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

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

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

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

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

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

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

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

We now require the M^2 **canonical correlations** between the columns in $\mathbf{D}$ and those in $\mathbf{E}$. To continue, we will digress briefly to define the canonical correlations. Let $\mathbf{d}_1^*$ denote a linear combination of the columns of $\mathbf{D}$, and let $\mathbf{e}_1^*$ denote the same from $\mathbf{E}$. We wish to choose these two linear combinations so as to maximize the correlation between them. This pair of variables are the first canonical variates, and their correlation r_1^* is the first canonical correlation. In the setting of cointegration, this computation has some intuitive appeal. Now, with $\mathbf{d}_1^*$ and $\mathbf{e}_1^*$ in hand, we seek a second pair of variables $\mathbf{d}_2^*$ and $\mathbf{e}_2^*$ to maximize *their* correlation, subject to the constraint that this second variable in each pair be orthogonal to the first. This procedure continues for all M pairs of variables. It turns out that the computation of all these is quite simple. We will not need to compute the coefficient vectors for the linear combinations. The squared canonical correlations are simply the ordered characteristic roots of the matrix,

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

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

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

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

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

21.3.4 ESTIMATING COINTEGRATION RELATIONSHIPS

Both of the testing procedures discussed earlier involve actually estimating the cointegrating vectors, so this additional section is actually superfluous. In the Engle and Granger framework, at a second step after the cointegration test, we can use the residuals from the static regression as an error correction term in a dynamic, first-difference regression, as shown in Section 21.3.2. One can then *test down* to find a satisfactory structure. In the Johansen test shown earlier, the characteristic vectors corresponding to the canonical correlations are the sample estimates of the cointegrating vectors. Once again, computation of an error correction model based on these first-step results is a natural next step. We will explore these in an application.

21.3.5 APPLICATION: GERMAN MONEY DEMAND

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

$$m_t - p_t = \mu + \beta y_t + \gamma i_t + \varepsilon_t, \quad \textbf{(21-8)}$$

where m_t, p_t, and y_t are the logs of nominal money demand, the price level, and output, and i is the nominal interest rate (not the log of). The equation involves trending variables (m_t, p_t, y_t), and one that we found earlier appears to be a random walk with drift (i_t). As such, the usual form of statistical inference for estimation of the income elasticity and interest semielasticity based on stationary data is likely to be misleading.

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

$$(m - p)_t - y_t = \mu + \gamma i_t + \varepsilon_t. \quad \textbf{(21-9)}$$

The left-hand side is the log of the inverse of the velocity of money, as suggested by Lucas (1988). An issue to be confronted in this specification is the exogeneity of the interest variable—exogeneity [in the Engle, Hendry, and Richard (1993) sense] of income is moot in the long-run equation as its coefficient is assumed (per Lucas) to equal one. Beyer explored this latter issue in the framework developed by Engle et al. (see Section 21.3.5).

TABLE 21.5 Augmented Dickey–Fuller Tests for Variables in the Beyer Model

Variable	m	Δm	$\Delta^2 m$	p	Δp	$\Delta^2 p$	$\Delta_4 p$	$\Delta\Delta_4 p$
Spec.	TS	RW	RW	TS	RW/D	RW	RW/D	RW
Lag	0	4	3	4	3	2	2	2
DF_τ	−1.82	−1.61	−6.87	−2.09	−2.14	−10.6	−2.66	−5.48
Crit.Value	−3.47	−1.95	−1.95	−3.47	−2.90	−1.95	−2.90	−1.95

Variable	y	Δy	RS	ΔRS	RL	ΔRL	$(m - p)$	$\Delta(m - p)$
Spec.	TS	RW/D	TS	RW	TS	RW	RW/D	RW/D
Lag	4	3	1	0	1	0	0	0
DF_τ	−1.83	−2.91	−2.33	−5.26	−2.40	−6.01	−1.65	−8.50
Crit.Value	−3.47	−2.90	−2.90	−1.95	−2.90	−1.95	−3.47	−2.90

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

$$(m - p)^* = \delta_0 + \delta_1 y + \delta_2 RS + \delta_3 RL + \delta_4 \Delta_4 p, \qquad \textbf{(21-10)}$$

where RS is a short-term interest rate, RL is a long-term interest rate, and $\Delta_4 p$ is the annual inflation rate—the data are quarterly. The first step is an examination of the data. Augmented Dickey–Fuller tests suggest that for these German data in this period, m_t and p_t are $I(2)$, while $(m_t - p_t)$, y_t, $\Delta_4 p_t$, RS_t, and RL_t are all $I(1)$. Some of Beyer's results which produced these conclusions are shown in Table 21.5. Note that although both m_t and p_t appear to be $I(2)$, their simple difference (linear combination) is $I(1)$, that is, integrated to a lower order. That produces the long-run specification given by (21-10). The Lucas specification is layered onto this to produce the model for the long-run velocity,

$$(m - p - y)^* = \delta_0^* + \delta_2^* RS + \delta_3^* RL + \delta_4^* \Delta_4 p. \qquad \textbf{(21-11)}$$

21.3.5.a COINTEGRATION ANALYSIS AND A LONG-RUN THEORETICAL MODEL

For (21-10) to be a valid model, there must be at least one cointegrating vector that transforms $\mathbf{z}_t = [(m_t - p_t), y_t, RS_t, RL_t, \Delta_4 p_t]$ to stationarity. The Johansen trace test described in Section 21.3.3 was applied to the VAR consisting of these five $I(1)$ variables. A lag length of two was chosen for the analysis. The results of the trace test are a bit ambiguous; the hypothesis that $r = 0$ is rejected, albeit not strongly (sample value $= 90.17$ against a 95% critical value $= 87.31$) while the hypothesis that $r \leq 1$ is not rejected (sample value $= 60.15$ against a 95% critical value of 62.99). (These borderline results follow from the result that Beyer's first three eigenvalues—canonical correlations in the trace test statistic—are nearly equal. Variation in the test statistic results from variation in the correlations.) On this basis, it is concluded that the cointegrating rank equals one. The unrestricted cointegrating vector for the equation with a time trend added, is found to be

$$(m - p) = 0.936y - 1.780\Delta_4 p + 1.601RS - 3.279RL + 0.002t. \qquad \textbf{(21-12)}$$

(These are the coefficients from the first characteristic vector of the canonical correlation analysis in the Johansen computations detailed in Section 21.3.3.) An exogeneity test—we have not developed this in detail; see Beyer (1998, p. 59), Hendry and Ericsson (1991), and Engle and Hendry (1993)—confirms weak exogeneity of all four right-hand-side variables in this specification. The final specification test is for the Lucas formulation and elimination of the time trend, both of which are found to pass, producing the cointegration vector,

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

The conclusion drawn from the cointegration analysis is that a single-equation model for the long-run money demand is appropriate and a valid way to proceed. A last step before this analysis is a series of Granger causality tests for feedback between changes in the money stock and the four right-hand-side variables in (21-12) (not including the trend). The test results are generally favorable, with some mixed results for exogeneity of GDP.

21.3.5.b TESTING FOR MODEL INSTABILITY

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

$$f(\mathbf{z}_t | \mathbf{z}^0_{t-1}, \Psi) = f[(m-p)_t | y_t, \Delta_4 p_t, RS_t, RL_t, \mathbf{z}^0_{t-1}, \boldsymbol{\Psi}_1] \times g(y_t, \Delta_4 p_t, RS_t, RL_t | \mathbf{z}^0_{t-1}, \boldsymbol{\Psi}_2).$$

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

$$\begin{aligned}\Delta(m - p)_t = &\sum_{i=1}^{4} c_i \Delta(m-p)_{t-i} + \sum_{i=0}^{4} d_{1,i}\Delta(\Delta_{4p_{t-i}}) + \sum_{i=0}^{4} d_{2,i}\Delta y_{t-i} \\ &+ \sum_{i=0}^{4} d_{3,i}\Delta RS_{t-i} + \sum_{i=0}^{4} d_{4,i}\Delta RL_{t-i} + \lambda(m-p-y)_{t-1} \\ &+ \gamma_1 RS_{t-1} + \gamma_2 RL_{t-1} + \mathbf{d}_t'\boldsymbol{\phi} + \omega_t, \qquad \textbf{(21-13)}\end{aligned}$$

where $\mathbf{d}_t$ is the set of additional variables, including the constant and five one-period dummy variables that single out specific events such as a currency crisis in September, 1992.[16] The model is estimated by least squares, "stepwise simplified and reparameterized." (The number of parameters in the equation is reduced from 32 to 15.[17])

[16] Beyer (1998, p. 62, footnote 4).

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

The estimated form of (21-13) is an autoregressive distributed lag model. We proceed to use the model to solve for the long-run, steady-state growth path of the real money stock, (21-10). The annual growth rates $\Delta_4 m = g_m$, $\Delta_4 p = g_p$, $\Delta_4 y = g_y$ and (assumed) $\Delta_4 RS = g_{RS} = \Delta_4 RL = g_{RL} = 0$ are used for the solution[18]

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

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

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

Analysis then proceeds based on this estimated long-run relationship.

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

21.4 NONSTATIONARY PANEL DATA

In Section 11.10, we began to examine panel data settings in which T, the number of observations in each group (e.g., country), became large as well as n. Applications include cross-country studies of growth using the Penn World Tables,[19] studies of purchasing power parity,[20] and analyses of health care expenditures.[21] In the small T cases of longitudinal, microeconomic data sets, the time-series properties of the data are a side issue that is usually of little interest. But when T is growing at essentially the same rate as n, for example, in the cross-country studies, these properties become a central focus of the analysis.

The large T, large n case presents several complications for the analyst. In the longitudinal analysis, pooling of the data is usually a given, although we developed several extensions of the models to accommodate parameter heterogeneity (see Section 11.10). In a long-term cross-country model, any type of pooling would be especially suspect. The time series are long, so this would seem to suggest that the appropriate modeling strategy would be simply to analyze each country separately. But this would neglect the hypothesized commonalities across countries such as a (proposed) common growth rate. Thus, the time-series panel data literature seeks to reconcile these opposing features of the data.

As in the single time-series cases examined earlier in this chapter, long-term aggregate series are usually nonstationary, which calls conventional methods (such as

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

[19]Im, Pesaran, and Shin (2003) and Sala-i-Martin (1996).

[20]Pedroni (2001).

[21]McCoskey and Selden (1998).

those in Section 11.10) into question. A focus of the recent literature, for example, is on testing for unit roots in an analog to the platform for the augmented Dickey–Fuller tests (Section 21.2),

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{m=1}^{L_i} \gamma_{im} \Delta y_{i,t-m} + \alpha_i + \beta_i t + \varepsilon_{it}.$$

Different formulations of this model have been analyzed, for example, by Levin, Lin, and Chu (2002), who assume $\rho_i = \rho$; Im, Pesaran, and Shin (2003), who relax that restriction; and Breitung (2000), who considers various mixtures of the cases. An extension of the KPSS test in Section 21.2.5 that is particularly simple to compute is Hadri's (2000) LM statistic,

$$\text{LM} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{t=1}^{T} E_{it}^2}{T^2\hat{\sigma}_\varepsilon^2}\right) = \frac{\sum_{i=1}^{n} KPSS_i}{n}.$$

This is the sample average of the KPSS statistics for the n countries. Note that it includes two assumptions: that the countries are independent and that there is a common σ_ε^2 for all countries. An alternative is suggested that allows σ_ε^2 to vary across countries.

As it stands, the preceding model would suggest that separate analyses for each country would be appropriate. An issue to consider, then, would be how to combine, if possible, the separate results in some optimal fashion. Maddala and Wu (1999), for example, suggested a "Fisher-type" chi-squared test based on $P = -2\Sigma_i \ln p_i$, where p_i is the p value from the individual tests. Under the null hypothesis that ρ_i equals zero, the limiting distribution is chi squared with $2n$ degrees of freedom.

Analysis of cointegration, and models of cointegrated series in the panel data setting, parallel the single time-series case, but also differ in a crucial respect.[22] Whereas in the single time-series case, the analysis of cointegration focuses on the long-run relationships between, say, x_t and z_t for two variables for the same country, in the panel data setting, say, in the analysis of exchange rates, inflation, purchasing power parity or international R & D spillovers, interest may focus on a long-run relationship between x_{it} and x_{mt} for two different countries (or n countries). This substantially complicates the analyses. It is also well beyond the scope of this text. Extensive surveys of these issues may be found in Baltagi (2005, Chapter 12) and Smith (2000).

21.5 SUMMARY AND CONCLUSIONS

This chapter has completed our survey of techniques for the analysis of time-series data. Most of the results in this chapter focus on the internal structure of the individual time series themselves. While the empirical distinction between, say, AR(p) and MA(q) series may seem ad hoc, the Wold decomposition theorem assures that with enough care, a variety of models can be used to analyze a time series. This chapter described what is arguably the fundamental tool of modern macroeconometrics: the tests for nonstationarity. Contemporary econometric analysis of macroeconomic data has added considerable structure and formality to trending variables, which are more

[22]See, for example, Kao (1999), McCoskey and Kao (1999), and Pedroni (2000, 2004).

common than not in that setting. The variants of the Dickey–Fuller and KPSS tests for unit roots are indispensable tools for the analyst of time-series data. Section 21.4 then considered the subject of cointegration. This modeling framework is a distinct extension of the regression modeling where this discussion began. Cointegrated relationships and equilibrium relationships form the basis of the time-series counterpart to regression relationships. But, in this case, it is not the conditional mean as such that is of interest. Here, both the long-run equilibrium and short-run relationships around trends are of interest and are studied in the data.

Key Terms and Concepts

- Augmented Dickey–Fuller test
- Autoregressive integrated moving-average (ARIMA) process
- Bounds test
- Canonical correlation
- Cointegrated
- Cointegration
- Cointegration rank
- Cointegrating vector
- Common trend
- Data-generating process (DGP)
- Dickey–Fuller test
- Equilibrium error
- Integrated of order one
- Nonstationary process
- Phillips–Perron test
- Random walk
- Random walk with drift
- Superconsistent
- Trend stationary process
- Unit root

Exercise

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

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

Applications

1. Using the macroeconomic data in Appendix Table F5.2, estimate by least squares the parameters of the model $c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \beta_3 c_{t-2} + \varepsilon_t$, where c_t is the log of real consumption and y_t is the log of real disposable income.
   a. Use the Breusch and Pagan LM test to examine the residuals for autocorrelation.
   b. Is the estimated equation stable? What is the characteristic equation for the autoregressive part of this model? What are the roots of the characteristic equation, using your estimated parameters?
   c. What is your implied estimate of the short-run (impact) multiplier for change in y_t on c_t? Compute the estimated long-run multiplier.
2. Carry out an ADF test for a unit root in the rate of inflation using the subset of the data in Appendix Table F5.2 since 1974.1. (This is the first quarter after the oil shock of 1973.)
3. Estimate the parameters of the model in Example 10.4 using two-stage least squares. Obtain the residuals from the two equations. Do these residuals appear to be white noise series? Based on your findings, what do you conclude about the specification of the model?